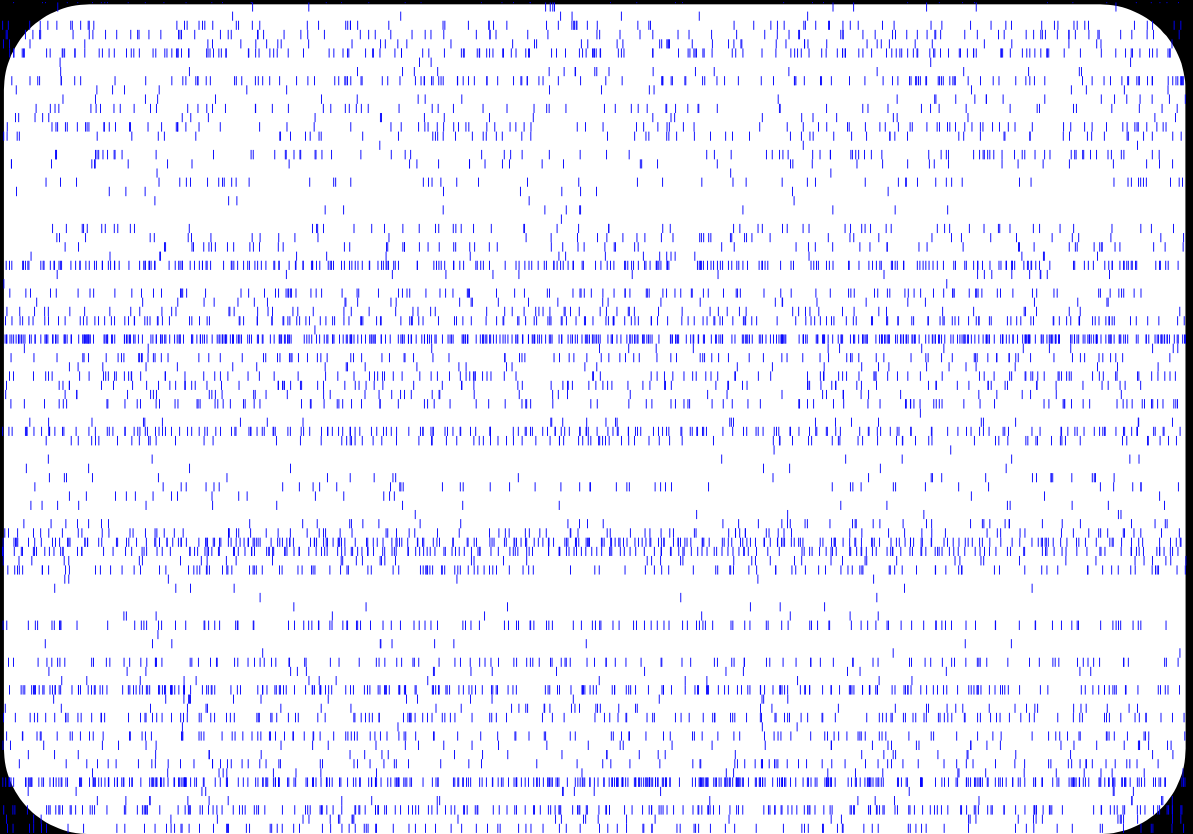


# Amplification in cortical networks

Jacques Edgard Angelo BOURG



Dissertation presented to obtain the Ph.D degree in Biology | Neuroscience  
Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,  
March, 2017



# Amplification in cortical circuits

Jacques Edgard Angelo BOURG

Dissertation presented to obtain the Ph.D degree in Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by:



**Champalimaud  
Foundation**

Supervisor : Alfonso Renart

Oeiras, March, 2017

**FCT** Fundação para a Ciência e a Tecnologia  
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

BD 51894/2012



UNIVERSIDADE  
**NOVA**  
DE LISBOA

*à ma famille  
à Camila*

Lawrence Krauss: -You often compare linguistics and some social sciences to physics. The virtue of physics is that it is simple. It is easy. Irrelevant details are easily removed, whereas as we talked about, you often bemoan that in linguistics and cognitive sciences, the ability to really make detailed progress is very difficult because all of the factors, and yet linguistics is therefore much less advanced than physics, and you said to me that that was an attraction, and I was surprised to hear about that.

Noam Chomsky: -It's an attraction, but I think what you describe is also true of early physics, so if you go back to the seventeenth century, Galileo had quite a hard time in convincing the funders, the aristocrats, that it made sense to study something that doesn't exist in nature, like a ball rolling down a friction-less plane. When they thought, if you study motion, why don't you study the growth of flowers ? -something that is real -or the way the leaves blow in the wind and so on and so forth. It took a long time for physics to get to the point where it became comprehended that if you want to understand the diversity and complexity of the phenomena of the world, you are going have to study highly idealised abstract models, even things which don't exist in the natural world, like friction-less plans, and then it's not like you throw away everything else, you hope that somehow you will get back to it, and that is the situation in the study of language in the 1950s [...].

Noam Chomsky and Laurence Krauss: An Origins Project Dialogue.

# Acknowledgements:

Quisiera agradecer a Alfonso por haberme dado esta oportunidad, por todo el apoyo y por todos estos años de trabajo duro, tan enriquecedores y productivos. A Leopoldo también quisiera darle las gracias por la calidad de nuestras interacciones. I would like to thank also my thesis committee Christian and Zach for being helpful and pragmatic, to the Professor Miguel Teixeira and to Gonzalo de Polavieja for having accepted to be part of my thesis jury. Also, I would like to acknowledge my opponents Guillaume and Michael for having accepted to be in part of my thesis, and because their work has been a source of many work and discussions. I would like also to thank all the members of the lab for all these fantastic years and to express them my esteem: Nivaldo, Katya, José, João, Raphael, Juan, Mafalda, Tor, Hanne, Roberto, André, Davide, Julien, Francisco and André. I would like to thank all the people that I had the pleasure to meet during these years: my INDP class, Nicolás, Fanny, Cristina, Daniel, Elvira, Julian, Jens, Simone, Vivek, Marta, Bahti, Adelina, Gabi and to Veronica. To Tor thanks for your craziness and your humanity. To Davide and Fanny, merci pour votre grande gentillesse, pour de très bons moments. A Caro, toutes nos années à l' INSA, tous les bons souvenirs et le privilège d'être ton pote. A Thomas, un grand merci pour ta générosité qui m'a rendu meilleur, pour une année au bâtiment D, et pour m'accompagner à Lisbonne. A Charles: merci pour nos conversations, pour partager une certaine vision des choses. A Irma, un gran cariño, nuestras conversaciones exigentes, un viernes en el barato. A Bertrand, merci pour tous les bons moments, les discussions, les bouffes, les cinés. A Thierry et Marion, un grand merci pour votre générosité, votre gentillesse et votre bonne humeur. A Tiago e a Nayara uma grande amizade, saudades de vós. A Camila todo o meu amor, obrigado pela tua força. Finalmente, je voudrais exprimer toute ma reconnaissance et mon affection à mes parents Nora et Jean-Christophe, à Amandine, à Daniel, à mes grands parents Jacques et Dominique, a Manuel, a Lu, a Helena, a mi abuela Nora.



# Abstract:

The brain is active all the time: it displays substantial spontaneous activity in awake and sleeping animals which doesn't receive sensory inputs. There are many open questions regarding this spontaneous activity: How is it generated? What is its function? What is its impact on sensory processing?

In this thesis we contribute to this problem by analysing data during the active -or desynchronized- brain state from urethane anesthetised rats. We find temporal structure at the population level in the form of unidimensional coherent fluctuations: alternatively, one half of the population decreases its firing rate while the other half increases it, keeping the population rate constant. We call this phenomenon competition, and we extensively characterise it.

Following, we ask the question: mechanistically, how might this competitive activity be generated? We attribute the intrinsic character of this competitive activity to the recurrent nature of the connectivity in the cortex. We revisit a known model of competition and we propose a new one which reproduces many observable dynamical quantities. In particular, our model uses a computational mechanism called non-normal amplification, from which we find signatures in the data.

In summary, in addition of revealing the competitive nature of the desynchronised state in the cortex, this study proposes a set of methodological as well as theoretical tools to analyse and model the relationship between connectivity and dynamics in neural circuits.

# Resumo:

O cérebro está ativo o tempo todo e exibe forte atividade espontânea tanto em animais acordados como em adormecidos em ausência de estímulos sensoriais. Há muitas questões abertas sobre esta atividade espontânea: Como é gerada? Qual é a sua função? Qual é o seu impacto no processamento sensorial?

A contribuição desta tese para essa problemática se encontra na análise de dados durante o estado ativo provenientes de ratos anestesiados com uretano. Encontramos uma estrutura temporal ao nível da população que toma a forma de flutuações temporais: alternativamente uma metade da população diminui a sua taxa de disparo enquanto a outra metade o aumenta, mantendo a taxa de disparo da população constante. Chamamos este fenômeno de competição, e o caracterizamos extensivamente

A etapa seguinte consistiu em pesquisar os mecanismos, que poderiam gerar esta atividade competitiva. Atribuímos o caráter intrínseco desta atividade competitiva à natureza recorrente da conectividade no córtex. Revisitamos um modelo conhecido da competição e propomos um novo modelo que reproduz várias quantidades dinâmicas observáveis. Em particular, o nosso modelo utiliza um mecanismo computacional chamado de amplificação não-normal, do qual encontramos assinaturas nos dados.

Em resumo, além de revelar a natureza competitiva do estado dessincronizado no córtex, este estudo propõe um conjunto de metodologias e de ferramentas teóricas para analisar e modelar a relação entre conectividade e dinâmica em circuitos neurais.

# Author contributions:

I also start this thesis by acknowledging the collaborators that took part in different phases of this project, and some of the people that took the time to discuss my work in depth. Excepting chapter 4, all this work was done under the guidance of Alfonso Renart.

- Chapter 2: Nivaldo Vasconcelos took part in the initial phase of the project, mainly establishing the competitive nature of the spontaneous activity in the active state. All the subsequent characterisations of the competitive state (reliability, time scale invariance, spatio-temporal description) was done solely with Alfonso Renart. The studied data set was collected by Liad Hollender and Péter Barthó in Rutgers University, U.S.A. We acknowledge Dmitry Kobak for discussions. Thanks to Michael Okun for sharing his code of population coupling. Thanks to Nuno Calaim, who implemented a matlab script to solve the Lyapunov equation. We thank Raphael Steinfeld, José Pardo Vazquez and João Afonso for sharing data.
- Chapter 3: We acknowledge Klaus Wimmer for having shared his code on the randomly connected network, along with Jaime de la Rocha and Albert Compte for discussions. We thank Guillaume Hennequin for discussions.
- Chapter 4: Nicolás Morgenstern and Leopoldo Petreanu designed the study, and N. Morgenstern did the experimental work. My contribution to this project was circumscribed to the modelling part.

This work was funded by a scholarship from the Fundação para a Ciência e a Tecnologia of the Portuguese ministry of education and science (BD 51894/2012), and by the Champalimaud Foundation.



# Table of contents

Table of contents	xv
List of Figures	xix
<b>1 Introduction: spontaneous activity in the brain</b>	<b>1</b>
1.1 The local cortical circuit . . . . .	2
1.1.1 Cellular composition and connectivity rules . . . . .	2
1.1.2 Basic laminar structure . . . . .	3
1.2 Spontaneous activity in the brain . . . . .	5
1.2.1 Active and inactive states . . . . .	6
1.3 Brain states and behaviour . . . . .	8
1.3.1 Behavioural correlates of spontaneous activity . . . . .	8
1.3.2 Cortical states and attention . . . . .	9
1.4 Neuromodulation . . . . .	11
1.4.1 Sleep-wake cycle . . . . .	11
1.4.2 Effects of acetylcholine and norepinephrine on the syn- chronisation level of cortical populations . . . . .	11
1.4.3 Use of anesthesia to study brain states . . . . .	13
1.5 Cortical states and evoked activity . . . . .	14
1.5.1 Sensory responses in different cortical states . . . . .	14
1.5.2 How could the brain distinguish spontaneous activity from evoked activity ? . . . . .	16
1.6 Possible roles of spontaneous activity . . . . .	19
1.6.1 Developmental role . . . . .	19
1.6.2 Memory consolidation . . . . .	20
1.6.3 Bayesian inference . . . . .	21

1.6.4	Sampling information from the environment in a way that conciliates resting and alertness. Encoding of this information in packets that propagate around the brain	23
1.7	Theoretical considerations about the mechanistic origin of the spontaneous activity	24
1.7.1	Spontaneous firing at single cell level	24
1.7.2	Spontaneous activity as a neural network property	26
1.7.3	Possible sources of correlations	32
1.7.4	Amplification in cortical circuits: spontaneous activity reflects idiosyncratic features of the network connectivity	33
1.8	Aim of the thesis	35
<b>2</b>	<b>Temporal competitive structure during the desynchronised state</b>	<b>37</b>
2.1	Introduction: temporal structure during spontaneous activity	38
2.2	Results: competitive activity revealed during the desynchronised state	40
2.2.1	Known features of the desynchronised state	40
2.2.2	Raw phenomenon: competitive dynamics	42
2.2.3	How stable are the PC directions across the recording ?	46
2.2.4	Do all neurons participate in the competition ?	48
2.2.5	Role of global fluctuations in competitive activity	50
2.2.6	Single cell characteristics	56
2.2.7	Temporal characterisation of the competition	60
2.2.8	Spatial characterisation of the competition	64
2.3	Discussion	70
2.3.1	Summary	70
2.3.2	Temporal invariance of the competition	70
2.3.3	Physiology, connectivity and competition	71
2.3.4	Functional significance of competitive dynamics during spontaneous activity	72
2.3.5	Low dimensional dynamics in cortex	78
2.4	Methods	79
2.4.1	Recordings and experimental procedures	79
2.4.2	Preprocessing the spike times into spike counts	80

2.4.3	Coherence methods for determining the relative position of the shanks . . . . .	81
2.4.4	Methods on Spectral Analysis . . . . .	84
2.4.5	Non-parametric statistical tests . . . . .	85
2.4.6	Necessity and sufficiency of a given subspace of activity . . . . .	88
<b>3</b>	<b>Modelling of competitive activity</b>	<b>93</b>
3.1	Introduction: dynamics in neural circuits . . . . .	96
3.1.1	Motivation . . . . .	96
3.1.2	Modelling choices . . . . .	96
3.1.3	Amplification in computational neuroscience . . . . .	99
3.1.4	Different mechanisms generating amplification in neural networks . . . . .	100
3.2	Study of the E-I randomly connected network . . . . .	106
3.2.1	Network architecture and simulation . . . . .	106
3.2.2	The eigenspectrum of the correlation matrix doesn't reveal competitive activity . . . . .	107
3.3	Circuits generating competitive activity . . . . .	109
3.3.1	The two dimensional linear excitatory and inhibitory network . . . . .	109
3.3.2	Normal competitive amplification (NCA) . . . . .	112
3.3.3	Transient competitive amplification (TCA) . . . . .	114
3.4	Model predictions and comparison with the data . . . . .	117
3.4.1	NCA and TCA generate negative correlations among $E_1$ and $E_2$ . . . . .	117
3.4.2	Asymmetry in the connectivity generate difference in the variances of $E_1$ and of $E_2$ . . . . .	118
3.4.3	Asymmetry in the connectivity generate difference in covariances between the two populations and the inhibitory neurons . . . . .	118
3.4.4	NCA crosscorrelogram is symmetric while TCA has a lag in the crosscorrelogram . . . . .	119
3.4.5	Angle between the PC1 and the uniform . . . . .	119
3.4.6	Comparison with data . . . . .	120
3.5	Generating competitive low dimensional dynamics in a high dimensional network . . . . .	123
3.5.1	Limitations of the low dimensional motif . . . . .	123

3.5.2	Generating an expanded high dimensional low-rank matrix . . . . .	124
3.5.3	Adding noise to the connectivity . . . . .	126
3.6	Predictions of the high dimensional model . . . . .	130
3.6.1	Correspondance between the HD and the LD version of the model . . . . .	130
3.6.2	Comparison between data and model . . . . .	130
3.7	Other models of competition . . . . .	133
3.7.1	Alternative low dimensional motif to TCA . . . . .	133
3.7.2	Translation-invariant connectivity matrices might generate competitive activity, but with different characteristics . . . . .	135
3.8	Summary and discussion . . . . .	138
3.8.1	Summary . . . . .	138
3.8.2	Relation with previous work . . . . .	139
3.8.3	Functional relevance of the asymmetry . . . . .	141
3.8.4	On the specificity of model predictions . . . . .	142
3.8.5	Model fitting . . . . .	142
3.8.6	More experiments to refine the predictions ? . . . . .	144
3.8.7	Other possible experiments to expand our understanding about the role of competition . . . . .	144
3.9	Methods . . . . .	146
3.9.1	Balanced randomly connected network simulation . . . . .	146
3.9.2	The Lyapunov equation . . . . .	146
3.9.3	Mean trajectory and cross-correlograms for linear dynamical systems . . . . .	147
3.9.4	Cross-correlograms of spike trains . . . . .	147
3.10	Appendix . . . . .	149
3.10.1	Two dimensional linear excitatory and inhibitory network . . . . .	149
3.10.2	Study of the correlation in the 2D linear network . . . . .	150
3.10.3	Rotating the covariance matrix from the Schur to the canonical basis . . . . .	152
3.10.4	Spectral decomposition . . . . .	153
3.10.5	Computing the eigenvectors of a connectivity matrix . . . . .	154
3.10.6	Schur decomposition of a 2*2 matrix . . . . .	156



3.10.7	Adding the deterministic and the stochastic high dimensional connectivity matrices . . . . .	157
3.10.8	Summary of the implementation of the high dimensional model . . . . .	163
3.10.9	Translation-invariant connectivity matrices . . . . .	167
<b>4</b>	<b>Cortical neurons integrate common inputs from sensory thalamus</b>	<b>177</b>
4.1	Introduction: the visual pathway, from the retina to the primary cortex . . . . .	179
4.1.1	The Hubel and Wiesel model . . . . .	179
4.1.2	Amplification of the thalamo-cortical projection . . .	182
4.1.3	Amplification and connectivity . . . . .	183
4.1.4	Aim of this chapter . . . . .	184
4.2	Experiments . . . . .	185
4.2.1	Preliminary experiments: determining the distribution and strength of dLGN axonal projections across cortical layers of V1 . . . . .	185
4.2.2	Probing connectivity between two neurons, measuring the covariations induced by long range projections . .	186
4.3	Modelling the covariance of the membrane potential of a couple of patched cells as a function of the proportion of shared inputs . . . . .	188
4.3.1	Modelling the correlation as a function of the stimulation power . . . . .	188
4.3.2	Modelling the mean correlation . . . . .	191
4.3.3	Differences in mean correlation reflect differences in the proportion of shared axons . . . . .	193
4.3.4	Adding variability about the number of recorded sites in one location . . . . .	197
4.3.5	Examining the consequences of variability in the post synaptic current amplitude . . . . .	198
4.4	Stimulating at only one power and counting the fraction of correlated locations . . . . .	201
4.4.1	Numerical simulation reproducing stimulation at constant power over a grid of points . . . . .	201
4.4.2	Experimental results . . . . .	203

4.5	Discussion . . . . .	206
4.5.1	Summary . . . . .	206
4.5.2	Modelling correlations . . . . .	206
4.5.3	Connectivity and function . . . . .	207
4.5.4	Importance of long range projections to understand cortical functions . . . . .	209
4.6	Appendix . . . . .	209
4.6.1	Mathematical proofs . . . . .	209
4.6.2	Numerical simulation . . . . .	211

## **5 Determination of the number of statistically significant principal components 213**

5.1	Introduction . . . . .	214
5.2	State of the art . . . . .	217
5.2.1	Parallel analysis . . . . .	218
5.2.2	MAP test . . . . .	221
5.2.3	Cross-validation methods . . . . .	222
5.3	Proposed method . . . . .	225
5.3.1	Motivations . . . . .	225
5.3.2	Preprocessing . . . . .	225
5.3.3	Shuffling the data: preserve the variance, destroy the covariance . . . . .	226
5.3.4	Use of this shuffle to determine the statistical signifi- cance of the principal components of the data . . . . .	227
5.3.5	Signal and noise . . . . .	229
5.3.6	Detailed procedure . . . . .	231
5.3.7	Explanation . . . . .	232
5.4	Comparing different methods . . . . .	236
5.4.1	Comparing methods with different data-bases . . . . .	236
5.4.2	Comparing methods with different artificially gener- ated data sets . . . . .	238
5.5	Determining the different number of statistically significant components in different experiments during desynchronised state . . . . .	241
5.6	Discussion and conclusions . . . . .	241
5.6.1	Results from the comparison with simulated data . . . . .	241

5.6.2	Results from the analysis of spiking data from desyn- chronised state . . . . .	242
5.6.3	Advantage of the preprocessing to evaluate non-stationarity of the data . . . . .	242
<b>6</b>	<b>General conclusion and discussion</b>	<b>245</b>
6.1	Main subject of this work . . . . .	246
6.2	Results . . . . .	247
6.2.1	Competitive activity during desynchronised state . .	247
6.2.2	Modelling of competitive activity . . . . .	247
6.2.3	Probing the relationship between local and long range connectivity . . . . .	248
6.2.4	Methodological contributions . . . . .	248
6.3	Discussion . . . . .	249
6.3.1	A new form of amplification: feed-forward amplification	249
6.3.2	Overall view on amplification . . . . .	251
	<b>Bibliography</b>	<b>252</b>



# List of Figures

1.1	Canonical cortical microcircuit . . . . .	3
1.2	Feed forward output of the cortical column . . . . .	5
1.3	Inactive state and active state . . . . .	8
1.4	The AIM model of brain-mind state control . . . . .	12
1.5	Modelling amplification in cortical circuits . . . . .	34
2.1	Desynchronised state in layer V of the cortex . . . . .	41
2.2	Competitive activity . . . . .	44
2.3	Competitive activity across experiments . . . . .	45
2.4	Stability of the first PC, halves . . . . .	46
2.5	Stability of the first PC, halves . . . . .	47
2.6	Significance of the loadings . . . . .	49
2.7	Correlation of the correlation matrix . . . . .	49
2.8	Necessity of the population rate fluctuations to explain competitive state . . . . .	50
2.9	Necessity and sufficiency of $PC_1$ and mean activity . . . . .	53
2.10	Relation between population coupling and PC1 . . . . .	55
2.11	Firing rate versus PC1 loadings . . . . .	57
2.12	The mean power of single cell firing correlates with loading . . . . .	59
2.13	Competition happens at many time scales . . . . .	61
2.14	Correlation matrices at different time scales . . . . .	62
2.15	Alignment of the different PCs at different time scales . . . . .	63
2.16	Alignment of the different PCs at different time scales, after having destroyed the correlations at smaller times scales . . . . .	63
2.17	Summary statistic of PC1 alignment . . . . .	64
2.18	Scenarios of the spatial arrangements of the competition . . . . .	65
2.19	Competition is local at the level of single shanks . . . . .	66
2.20	Quantifying the spatial heterogeneity of the competition . . . . .	67

2.21	Evolution of the the standard deviation of the loadings with the distance . . . . .	69
2.22	Luczak’s hypothesis of directions of spontaneous and evoked activity in neural space . . . . .	73
2.23	Mean population response to a one second pure tone averaged across trials . . . . .	74
2.24	Biphasic population response to a click in active state . . . .	76
2.25	The null space hypothesis of evoked activity during the desynchronised state . . . . .	77
2.26	Binning method . . . . .	81
2.27	Reordering the shanks . . . . .	83
2.28	Method to assess necessity and sufficiency of a given subspace in order to explain the variance in a data set. . . . .	91
3.1	Example of normal amplification . . . . .	101
3.2	Example of contractive and non-contractive dynamics . . . .	104
3.3	Schematic representation of a Schur decomposition . . . . .	105
3.4	Randomly connected balanced network . . . . .	107
3.5	Randomly connected networks doesn’t display competitive activity . . . . .	108
3.6	Anatomical and functional representation of a linear network	110
3.7	2D E-I network . . . . .	111
3.8	Anatomical and functional representation of normal competitive amplification . . . . .	116
3.9	Dynamical predictions from NCA and TCA . . . . .	121
3.10	Cross-correlograms in the data . . . . .	122
3.11	Example of three connectivity matrices with different complexity . . . . .	126
3.12	Schur interpretation of high dimensional linear networks . .	129
3.13	Comparing low dimensional model with high dimensional model of TCA . . . . .	130
3.14	Comparison between model and data: 1 . . . . .	131
3.15	Comparison between model and data: 2 . . . . .	132
3.16	Translation invariant connectivity matrices amplify subspaces of dimension two . . . . .	138

3.17	Example of a connectivity matrix with a gaussian profile that generates more variance on the first plus mode than in the uniform . . . . .	176
4.1	Hubel and Wiesel model of orientation selectivity . . . . .	181
4.2	Contribution of the recurrent cortical connectivity to the amplification of thalamic inputs . . . . .	183
4.3	Method to relate local and long range connectivity . . . . .	187
4.4	Schematic of the model . . . . .	189
4.5	Variance as a function of the stimulation power . . . . .	190
4.6	Averaging correlations across stimulation sites with different biophysical characteristics . . . . .	192
4.7	Approximation relating the maximal mean correlation with the fraction of shared axons . . . . .	196
4.8	Similarity of two models of correlation . . . . .	198
4.9	Simulation of the fraction of correlated locations as a function of the mean number of recruited axon . . . . .	202
4.10	Fraction of correlated locations as a function of the mean number of inputs . . . . .	204
4.11	Measured number of correlated locations as function of the mean number of inputs . . . . .	205
5.1	Description of PCA and comparison with FA . . . . .	216
5.2	Effect of the sample size on the distribution of the eigenvalues . . . . .	219
5.3	Example of parallel analysis applied to real data . . . . .	220
5.4	Shuffling the data in order to determine the statistical significance of the first principal component . . . . .	228
5.5	Subtraction a subspace to see the statistical significance of the variance in the remaining subspaces . . . . .	230
5.6	Method summary . . . . .	234
5.7	Method iterations . . . . .	235
5.8	Correlation matrices used to generate data . . . . .	238
6.1	Feed-forward amplification . . . . .	250





# Chapter 1

## Introduction: spontaneous activity in the brain

HIGHLIGHTS

- In this first chapter we do a literature review concerning the spontaneous activity in the brain: its origin, its possible roles, its influence on perception and on behaviour.
- We present the distinction between asynchronous and synchronous population states, which are specific types of spontaneous activity observed in the cortex.
- We motivate why studying spontaneous activity might help us understand something about the cortical connectivity.

### Contents

1.1	The local cortical circuit . . . . .	2
1.2	Spontaneous activity in the brain . . . . .	5
1.3	Brain states and behaviour . . . . .	8
1.4	Neuromodulation . . . . .	11
1.5	Cortical states and evoked activity . . . . .	14
1.6	Possible roles of spontaneous activity . . . . .	19

1.7	Theoretical considerations about the mechanistic origin of the spontaneous activity . . . . .	24
1.8	Aim of the thesis . . . . .	35

---

## 1.1 The local cortical circuit

### 1.1.1 Cellular composition and connectivity rules

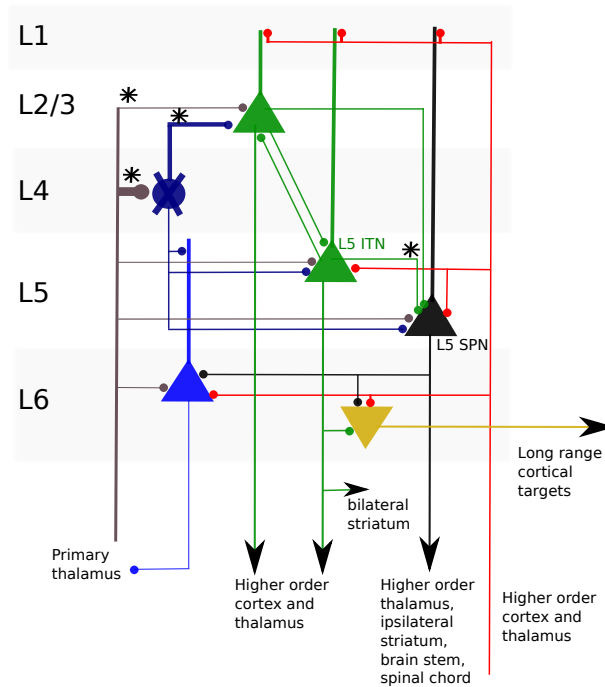
The cortex is an external thin layer of some millimetres that enfolds the mammal's brains. It is connected to sub-cortical structures such as the thalamus and the basal ganglia. Cortical neurons are of two major classes: **principal cells** and **inter-neurons**. Principal cells are excitatory, they express glutamate and they constitute around 80% of cortical neurons in rodents. The restant 20% of cortical cells are the **inter-neurons**: they express GABA, which tends to have an inhibitory effect on the post-synaptic membrane potential. There are neuron subclasses with particular genotypical and physiological differences. Concerning the principal cells, there are: intratelencephalic neurons, pyramidal tract neurons and corticothalamic neurons [60], [143]. Concerning the interneurons, there are somatostatin-expressing interneurons, parvalbumin-expressing interneurons and 5HT3A-receptor-expressing interneurons [60].

Neurons tend to be connected depending on the types of pre and post-synaptic neurons. There tends to be a over-representation of bidirectional connections between principal cells. These cells form interdigitated subnetworks in which similarly stimulus-tuned neurons tend to be preferentially connected [104]. However, this pattern of connectivity is not strict: not all neurons that respond to similar features are connected, and conversely not all connected neurons have similar stimulus preferences [59]. In the fourth chapter, we will see an example of spatial selectivity in the projections: the axons from the thalamus tend to target preferentially couples of cells that are connected between themselves. The preferential connectivity might well promote amplification of the cortical responses - in the sense of a multiplicative modulation of the tuning curve, as Olsen and colleagues have observed in L6 [110]. For somatostatin and parvalbumin-expressing interneurons, the connection probability with a neighboring principal cell is close to 100%,

which means that the connection is non specific, and that they fire to a broad class of stimuli .

### 1.1.2 Basic laminar structure

The connections across cortical areas (sensory, motor, associative) seem to be very stereotypical and follow a basic pattern called a **canonical micro-circuit**.



*Figure 1.1: Canonical cortical microcircuit. In this sketch we only show the principal cells, and we ignore the fact that inside each layer, pyramidal cells tend to connect with cells of the same subclass. This figure was adapted from [59].*

There are quantitative differences of this circuit across areas and species [60], but nonetheless, this motif suggests the existence of a basic computational module that is the substrate of the learning and/or the execution

of a full spectrum of high-order functions like fine sensory discrimination, speech, decision making or motor skills. Many people think that the canonical microcircuit must play a basic computational role, like the transistor in digital electronics. The transistor is the building block of logical gates (and, or, nand, ...). These logical gates are a very powerful tool, because every Boolean function -correspondence between digital inputs and outputs- is implementable as a composition of logical gates. In a subsequent chapter we will show an example of a model of a recurrent network that intends to model the fifth layer of the cortex. In this model, when we vary parametrically the strength of the connections, we can show theoretically that there is a qualitative difference in way the network operates. Functionally, the first three **supragranular** layers L1,L2,L3, are the main origin and termination of intra-cortical connections within areas of the same hemisphere or to the opposite hemisphere, through the corpus callosum. The fourth layer (L4) is the main circuit input from the thalamus. In the chapter four, we will expand the discussion about the main thalamocortical projections into the cortex, that synapses mainly onto layer 4 but also in layer 2/3. From layer 4, the information then flows in a feed-forward way to L2/3 and from L2/3 to L5.

Determining the connectivity rules within and across layers is a major challenge that might reveal the computations that are taking place in this circuit. For example Kampa and colleagues [76], recorded simultaneously triplets of neurons, placing one patching electrode in L2/3 and the two others in L5. In a second experiment, they also placed two electrodes in L2/3 and one in L5. By injecting current in one of the three electrodes, it is possible to observe whether they elicit excitatory post-synaptic potentials in the two other electrodes and then obtain a statistical view of the way neurons tend to connect (figure 1.2 left panel). They conclude that pairs of neurons in layer five tend to receive more shared input from cells in L2/3, when they are connected. This suggest that there are sub-networks in layer five that tend to receive specific shared inputs from L2/3 (see figure 1.2), which in turn receive specific inputs from L4, as shown by Callaway's group in 2005 [164].

In figure 1.1, we marked with an asterisk the main connections that we will consider in this thesis. We will first analyse data from primary somatosensory and auditory cortex in L5. Layer 5 is a major output projection of the cortex. As we see there are two cell subclasses in layer 5 (intratelen-

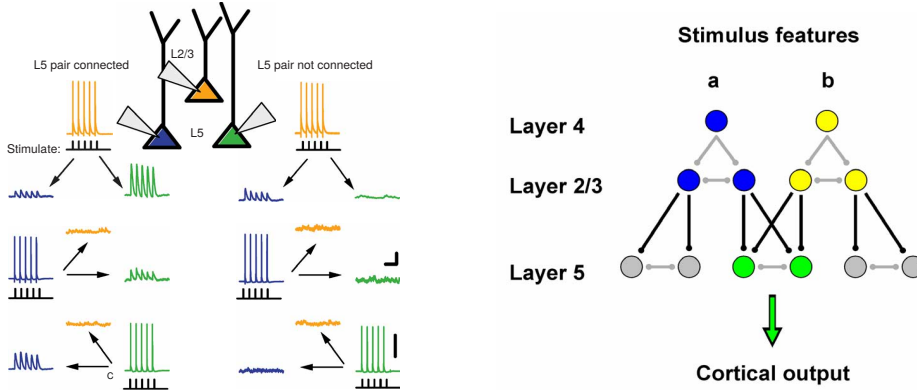


Figure 1.2: Feed forward output of the cortical column. Left panel: experiment to map the connectivity between L2/3 and L5. Right panel: schema the spatial connectivity of the feed forward network going from L4 through L2/3 to L5. Adapted from [76].

cephalic neurons and sub-cerebral projection neurons), which connect asymmetrically, the projections going mainly from intratelencephalic neurons to sub-cerebral projection neurons. However, intratelencephalic neurons tend to have a larger connection probability between themselves with respect to sub-cerebral projection neurons [81]. Also, intratelencephalic neurons tend to show moderate firing rate, whereas sub-cerebral projection neurons have a periodic spatial organisation and can spike in bursts. Finally sub-cerebral projection neurons project to the striatum, the brain stem, and to the spinal chord.

As we mentioned earlier, the whole picture of how the canonical microcircuit works is more complex, because we need to add the inter-neurons, which are inhibitory. For example, it has been reported that pyramidal neurons in L6 have a net inhibitory effect on the other cortical layers [110].

## 1.2 Spontaneous activity in the brain

We refer to **spontaneous activity** as the brain activity in the absence of an apparent sensory input or motor output. It is also termed resting-state activity, by opposition to **evoked activity**.

The first reports of spontaneous activity in the brain go back to 1891 when Adolf Beck placed electrodes on the surface of the brain of rabbits and dogs, and observed fluctuating activity. In turn, in 1924, Hans Berger placed electrodes on the scalp of human subjects, and recorded the first human electroencephalogram -**EEG**-. Using the EEG, Berger reported the existence of alpha waves (7.812 -13.28 Hz) when the subject closed his eyes, and their replacement by faster beta waves (12.5 - 30 Hz), when the subject opened the eyes. Also, he reported alterations in the EEG during epileptic seizures. The EEG was important for the discovery that during sleep, the surface of the brain went through different dynamical regimes, also called cortical states.

With the arrival of multi-unit extracellular recordings in the 2000's, the neural correlates of the cortical states at the population level started to be explored in animal models.

In addition to the cortex, spontaneous activity has been observed in the retina, in the cochlea, in the spinal cord, in the cerebellum, in the thalamus, in the basal ganglia and in the hippocampus [16],[61]. Aside from the introductory remarks in this chapter, in subsequent chapters we will focus on spontaneous activity in the cortex.

### 1.2.1 Active and inactive states

Active and inactive states are dynamical properties of a cortical networks, which are known to be recurrently connected. The state qualifies the macroscopic behaviour of the network, which results from the way neurons fire in relation to each other.

**Inactive state** designates a regime in which there is strong synchronisation of all the neurons: during the **up phases**, all neurons fire together, and during the **down phases**, all neurons are silent, and this is also visible at the level of the multi-unit activity (**MUA**), or **population firing rate**. This alternation between up and down phases is strictly speaking not an oscillation, in the sense that the transitions between up and down states are not periodic in time, they seem random, but we will use anyway this terminology in a sloppy way. The slow oscillation has a power spectrum that is concentrated in a frequency band from 0.5 to 4 Hz. Also, because of this sharp transitions between activity and silence, pairs of cells tend to covary together around their respective mean, so that on average, the mean

pairwise correlation - the average of the correlations between all pairs of recorded neurons- is positive (see figure 1.3). Other synonyms of inactive state that are also present in the literature and that we will commonly use are synchronised, **inactivated** or **deactivated**.

At the single cell level, the membrane potential makes slow-frequency, large-amplitude changes, that correlate with the up and down phases [30], [131], [142]. The slow oscillation propagates around the different cellular layers of the cortex [131], and across cortical columns and brain areas [61] .

The **active** state (also called **activated** or **desynchronised state**), presents on the contrary a more tonic firing rate of the cells, and a constant MUA (see figure 1.3). The histogram of pairwise correlations is centered around zero, with a small but positive mean, so that half of the pairs of cells are positively correlated and half are negatively correlated. At the single cell level, the membrane potential is depolarised close to the firing threshold [166].

In the computational literature that aims at modelling these two brain states there are two technical terms: **synchronous**, and **asynchronous**, which have a precise definition, but whose characterisation might be more or less close to the real cortical states.

Both synchronised and desynchronised states can be identified using the local field potential **LFP**, which is the extracellular membrane potential generated by the transmembrane currents: in active state the LFP displays little power in the lower frequencies (delta) and high power in the higher frequencies, whereas in inactive state, it follows the alternation between up and down phase, but in an inverted way with respect to the MUA, because the reference is placed extracellularly. Finally, what we call active and inactive state are in reality extremes of a continuum of desynchronisation: in certain situations the cortex might be more or less locked to the extremums, and in other situations to have a time varying desynchronisation level.

There is a tight correspondence between the activity in the cortex and the activity in the hippocampus (even if the dynamics in the hippocampus are different). The equivalent of the synchronous state in cortex is called large irregular activity in the hippocampus (**LIA**), and the equivalent of the desynchronised activity is called **hippocampal theta**.

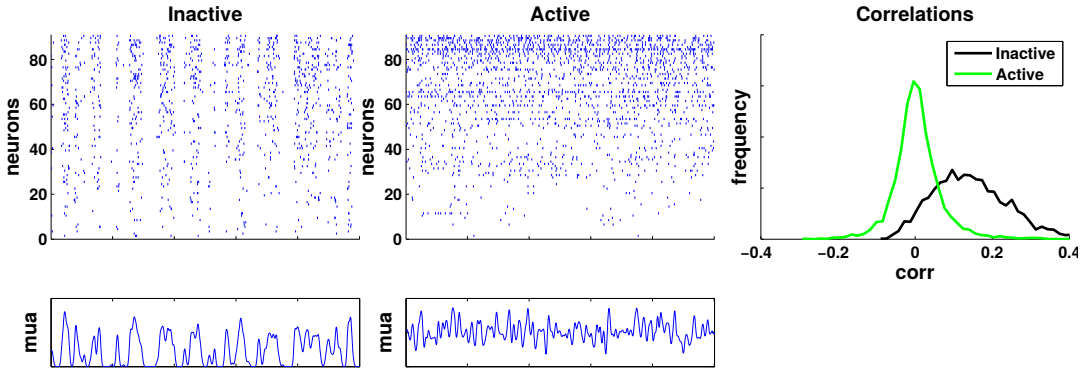


Figure 1.3: Inactive state and active state. The population rasters present 5s of data recorded from layer V in primary auditory cortex. Every dot represents an action potential.

## 1.3 Brain states and behaviour

### 1.3.1 Behavioural correlates of spontaneous activity

During sleep, the distinction between cortical states is clear because the difference in the level of desynchronisation between rapid eye movement -REM- and of synchronisation during slow wave sleep -SWS- is very pronounced. In SWS the eyes execute rolling movements, whereas in REM sleep the eyes move in random directions.

Both in active as in inactive state, changes in the pupil size correlate very well both with the slow "oscillation" of the synchronous activity also with important changes in the desynchronisation level [98].

In awake state, the cortex wanders between inactive and active state [98]. However, there is still not a complete taxonomy of fine behavioural correlates of both states in awake state.

The effort of relating behaviour with the brain rhythms started in the seventies with the pioneering work of C.H Vanderwolf, that noted the correlation of the hippocampal LFP with the ongoing behaviour of the rat [150]: "Trains of rhythmical 6-12 c/sec [hippocampal theta] waves in the hippocampus and medial thalamus precede and accompany gross **voluntary types of movement** such as walking, rearing, jumping, etc. Behavioural immobility



(in the alert state) and **automatic movement patterns** such as blinking, scratching, washing the face, licking or biting the fur, chewing food or lapping water are associated with irregular hippocampal activity [LIA]". We now know that theta waves in the hippocampus correlate with desynchronisation in the cortex and LIA correlates with synchronisation.

Futhermore, Poulet and Petersen [119] showed that during these periods of quiet wakefulness, when the cortex was more synchronised, the membrane potential of nearby cells was also highly correlated.

Zagha and colleagues [166] observed recently in 2015 that the mice's deep layers of motor cortex could stay activated for long periods, even in the case of absence of stimuli or of observable movements. This happened while the animals were achieving good performance in a behavioural paradigm that consisted in withholding from licking and licking after a whisker touch. Interestingly, towards the end of the behavioural session, when the mice were tired and did more mistakes, motor cortex was more desynchronised, in particular before miss trials in comparison to hit trials.

Finer, non subjective quantification of behaviour might help understanding the subtleties of the correlation between brain states and awake behaviour, which will require techniques like video segmentation and machine learning to find regularities between the levels of synchronisation and the behaviour in freely moving conditions. A more controlled approach consists in using head-fixed mice on top of styrofoam balls and closed loop virtual reality environments [62], [107].

### 1.3.2 Cortical states and attention

**Attention** refers to the process by which "organisms select a subset of available information upon which to focus for enhanced processing" [156]. Some examples of attention are: orienting the head towards a strong unexpected sound, focusing on a conversation and do not perceive that someone steals your wallet, or searching for a contact lens on the floor. The first example is an example of **bottom-up attention** whereas the other two are examples of **top-down attention**, also called **selective attention**. The difference between top-down and bottom-up attention depends then on the way of filtering out the information: whether it is provoked by the saliency

of the stimulus characteristics (bottom-up), or whether an instructive signal sharpens our senses to detect and process a particular kind of stimulus at the expense of others. Here we will concentrate in top-down attention effects.

A classic behavioural paradigm to study attention in monkeys was proposed by Moran and Desimone in 1985 [103]. A monkey fixates the gaze in a fixation point center, and using its visual periphery it has to solve the task to obtain water. Certain cells in the visual system possess what is called a **receptive field**: they spike when a particular type of stimulus is present in a given region of the visual space. In this experiment, Moran and Desimone recorded cells which spiked when horizontal and vertical red bars -but not green bars-, were presented on their receptive field. In their experiment, both a red bar and a green bar were presented in the receptive field of the recorded cell. By blocks, to solve the trial, the monkey has to pay attention to only one of the two bars, the red or the green bars. To win a drop, the animal has to quickly press a bar when it sees the same orientation in the sample and in the test, otherwise there is no reward. This kind of task is called a **match-to-sample task**, and by design, the subject needs to pay attention twice in order to be solved. If the response of the recorded cell was purely sensory -and not modulated by attention -, the cell should spike in a similar manner in all conditions. However, what Moran and Desimone found, is that in  $V_4$  (fourth cortical visual area) and in **IT** -inferior temporal cortex, but not in  $V_1$ , there are cells which show modulation of attention: when the effective stimulus (the red bars) was attended either during the sample or during the test, the recorded cell had a strong response, but when the monkey attended the irrelevant stimulus (green bars), the cell was almost silent, in spite of the presence of the relevant stimulus in the receptive field.

Using a similar behavioural paradigm as Moran [103], Fries and colleagues showed that attention reduced low frequency synchronisation of  $V_4$  neurons that represented the effective stimulus, and importantly, that the observed desynchronisation was local in space. They also showed that attention comes with an increase in **gamma frequency** (35-90 Hz). Finally, Mitchell and colleagues [99], as well as Cohen and Maunsell [26] showed in 2009 that attention reduces variability, noise correlations and low-frequency fluctuations.

Both attention and cortex desynchronisation have many characteristics in common: reduced power in low frequencies of the LFP, reduced trial to

trial variability, and reduced noise correlations [61]. Also, the neuromodulatory systems involved in synchronisation, that we will mention in the next section, are also involved in attention. However because the neuromodulatory systems broadcast signals in a more global way at the level of the cortex, attention involves supplementary mechanisms in order to gain spatial selectivity. Deco and Thiele [34] have proposed, using biophysical simulations, that cholinergic feedback projections from higher cortical areas might mediate local attentional modulation of cortical circuits.

## 1.4 Neuromodulation

### 1.4.1 Sleep-wake cycle

The chemical basis of cortical and hippocampal spontaneous activity are studied since the 70's (see [151]), and in particular the neuromodulators that are responsible for the alternations between sleep and awake and for the variations in brain state in awake and in sleep (REM vs non-REM). The neuromodulators that induce these variations are the noradrenergic system -situated in the locus coeruleus- the glutamatergic system -located in the thalamus- [61], [98] and the cholinergic system, located in the basal forebrain.

In the following figure 1.4, we see that during the sleep-wake cycle (circular arrows), J. Allan Hobson [66] shows the covariation of three physiological variables (activation, modulation and input-output gating). We can see that sleep correlates with low levels of norepinephrine, whereas the levels of acetylcholine might be both high during awake and sleep periods.

Xu and colleagues [163] perturbed optogenetically different cell types in the basal forebrain and were able to elicit wakefulness activating successively cholinergic, glutamatergic and parvalbumin-positive (PV+) GABAergic neurons. Reciprocally, they induced non-REM sleep, when activating somatostatin-positive (SOM+) GABAergic neurons.

### 1.4.2 Effects of acetylcholine and norepinephrine on the synchronisation level of cortical populations

Goard and Dan [51] showed in 2009, that brief stimulations of the nucleus basalis -a cholinergic nucleus in the basal forebrain- induced strong decor-

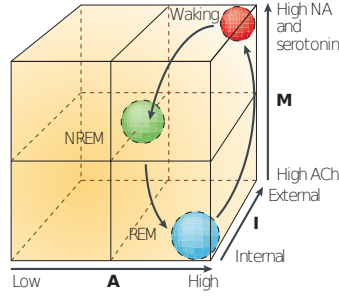


Figure 1.4: The AIM model of brain-mind state control [66]. *A* refers to activation, *M* to modulation and *I* to input-output gating, the process that controls the access to the sensory information from the external world and which prevents sending motor commands to the muscles.

relation between neurons in the visual cortex of rats. The acetylcholine released by the basal forebrain acts on the interneurons **disinhibiting** the pyramidal cells in layer 5. Precisely, the fifth layer of the cortex is the place in which the up and down phases are generated to then spread out to the other layers, as shown in slices by Sanchez-Vivez and McCormick [131].

Castro-Alamancos and Gulati [22] showed that the action of the cholinergic system and of the noradrenergic system led to different types of desynchronised states: in the case of cholinergic stimulation, the activated state maintained a constant population firing rate with respect to the inactivated state, whereas the norepinephrinergic cortical stimulation decreased the overall firing rate.

More recently, in 2013, Polack et al. [118] performed whole cell recordings in head-fixed mice on a spherical treadmill. By separating the membrane potentials of single cells according to whether the mice were running or resting, they computed the distributions of the membrane potentials in both conditions. They found that both during immobility and locomotion, the distribution of membrane potentials was unimodal. However, during locomotion the distribution had a larger mean membrane potential, and was sharper, meaning that the variance was smaller. Thereby, they proceeded to apply local perfusions of agonists and antagonists of both drugs, and concluded that during the resting periods, the cholinergic input was responsible for maintaining unimodal the shape of the distribution of membrane poten-

tials. In turn, norepinephrine was responsible for the depolarisation of the mean membrane potential during locomotion.

In summary, these results show a complex and unfinished picture of what brain states are, and of how they are regulated by the neuromodulatory systems. The picture of an unidimensional continuum between synchronised and desynchronised cortical states may be much less clear and higher dimensional than we usually think.

### 1.4.3 Use of anesthesia to study brain states

The use of spherical treadmills and head fixed rodents is very recent. For practical reasons, instead of recording neural activity directly from behaving animals, the traditional approach has been of recording from anaesthetised animals.

The first kind of anaesthetics that were used to study the brain in the 60's and in the 70's were barbiturates like thiopental sodium. However, this drug has a strongly depressing effect and doesn't induce population dynamics that resemble the ones observed during sleep or awake. In vitro, barbiturates hyperpolarise the membrane potentials, only allowing some rare synaptic events of small amplitude [35]. In vivo, the barbiturates tend to suppress the effects of the recurrent connectivity on the dynamics, letting only the stimulus driven feed-forward component of the currents elicit activity. This fact was not really understood at the time, but didn't prevent a whole generation of neuroscientists like Hubel and Wiesel to successfully study sensory processing. When it became evident that ongoing activity was an interesting topic and that it could have an impact on sensory processing, researchers started trying other anaesthetics that don't preclude it.

Clement et al. [25] showed that it is possible to elicit periods of both high and low synchronisation using urethane anesthesia. In this study they show that state alternation is a cholinergic phenotype and it is independent on norepinephrine. The rate of these brain state alternations did not correlate with variations in the concentration level of urethane. The activity recorded during sleep and after the drug's injection was found to be very similar, both at the level of the physiological correlates (muscular tone, respiration rate, cardiac frequency) as well as in the LFP (Power, time intervals between states). This means that urethane anaesthetised animals seem a good preparation to study sleep and brain states.

Pachitariu and colleagues [111] also use alternative anaesthetics: they use ketamine/xylazine to lock the cortex in synchronous state and use fentanyl/medetomidine/midazolam to obtain desynchronised states.

## 1.5 Cortical states and evoked activity

### 1.5.1 Sensory responses in different cortical states

The study of spontaneous activity is inextricably linked with the study of sensory processing in the brain, because the neural representations result from an interaction of the feed-forward activity coming from the thalamus, with the ongoing activity that is widespread around the cortex. This interaction can be quite complex: Curto et al. [31] predict the population responses to clicks on a trial by trial basis, fitting data to a model which is non-linear in the synchronised states and approximately linear in the desynchronized states.

We will examine here the evidence that different cortical states elicit different neural responses in different sensory modalities, the way this impacts the fidelity of sensory representations and some of the mechanisms that provoke these differences in the responses across states.

Few years after the first discoveries of Hubel and Wiesel [72],[73], Wurtz replicated their experiments using awake trained monkeys [162]. He concluded that the basic organization of the receptive fields in striate cortex was similar to the organisation of the receptive fields in paralysed, anesthetized cats and monkeys measured by Hubel and Wiesel. Surprisingly, Wörgötter et al. [161] found in 1998 that the receptive fields of complex cells are wider in the visual space during synchronized states and smaller during non-synchronized states. Wörgötter reports that, for a 2.5 fold power increase in the range 1-4 Hz, the receptive field grows by 27 %, becoming less selective.

Recent findings [12], [107] point at increases in neural responsiveness as the cortex desynchronises during locomotion. In these studies, drifting gratings are presented to head fixed mice while the mice lay on a treadmill, and cells are recorded in the primary visual area ( $V_1$ ). These studies consistently report additive and multiplicative gains of the tuning curves in  $V_1$  during desynchronised state. The anatomical projection from motor cortex to auditory cortex that might be involved in these locomotion-associated changes of auditory cortical response was described by Schneider et al. [133].

Marguet and Harris [96] presented amplitude modulated noise stimuli to urethane anesthetised rats, and asked to which extent they could predict single neural activities based on the LFP or on the stimulus. During desynchronised activity, the neural activity could be predicted from both the stimulus and the LFP. For the synchronised activity, they found that the activity of individual neurons was strongly predictable from the LFP and poorly predictable from the AM noise envelope. Therefore, in presence of such AM noisy stimuli, the cortical activity is largely decoupled from the stimulus. This doesn't mean that during synchronised activity, the cortex can not be entrained by external stimuli: for more punctuate stimuli like clicks or tones, the cortex responds reliably transitioning from a down phase to an up phase [89], [31].

In auditory cortex of gerbils, Pachitaru and colleagues [111] probed the population response with other stimuli. Using pure tones, frequency modulated tones and speech, they concluded that the responsiveness, the selectivity, the reliability and the temporal precision of the population was higher during desynchronised state than during the synchronised state. Another relevant aspect of this study is that it examines how the representation capacity of auditory cortex is modified as a function of the states, i.e. to which extent different sounds evoke different population responses. The authors found that when presenting different speech sounds, the similarity between the responses was higher during the synchronised states than during the desynchronised state.

A **PSTH** or the peri-stimulus time histogram is the average response of a cell following the presentation of a stimulus. When plotting the PSTH of auditory cells at their best response frequency, Pachitaru and colleagues, find that if the brain could average out the slow oscillation in a single trial, it would find on average similar responses -not perfectly although- in both active and inactive states. However these responses are more variable in inactive states [12], [99].

Given the evidence from the literature, it seems that the information quality about the details of a stimulus conveyed by the neural populations is smaller in the inactive state with respect to the active state. Is the synchronised state during awake a watchful state of lower energy consumption, whose first role is not to encode properly all the details of a stimulus, but to facilitate the detection of stimulus onsets, so that the cortex desynchronises in case of danger for example?

Fanselow and Nicolelis do conclude in this direction in a study in somatosensory cortex [40]: "in the absence of whisker movements (quiet immobility), somatosensory neurons seem to be highly responsive to punctuate tactile stimuli but not to sequences of stimuli. In contrast, during the whisking state, when active exploratory whisker movements are used by the rat to gather tactile information, the temporal fidelity of sensory responses to rapidly presented stimuli is enhanced".

This difference in population response depending on the frequency of the stimuli is explained by Castro Alamancos [23] by the finding that the **rapid sensory adaptation** depends on the cortical state. Rapid adaptation is the decrease in the response of a neuron to a high frequency stimulus. We call it rapid, because it acts on a time scale of 100 ms. Castro Alamancos finds that during the desynchronised state, there is little adaptation whereas during synchronised state there is strong adaptation. The mechanisms advanced for the cortical adaptation are the depression of the thalamo-cortical synapses and an increase in inhibition.

As advanced by Castro Alamancos [23], the system might adapt its brain state and its response to stimuli to meet information processing demands dictated by behavioural contingencies. Luczak [89] recalls that the apparently noisy slow oscillation that might seem to degrade perception might in fact be the result of hidden variables that aren't under experimental control : "Only a small fraction of the input to a cortical column arises from primary sensory thalamus, and responses in sensory cortex can be affected by cognitive factors such as reward and attention, other sensory modalities, and ongoing oscillations".

### 1.5.2 How could the brain distinguish spontaneous activity from evoked activity ?

As we saw in previous sections, it seems that the brain state impacts the sensory representations in a way that seems to be adapted both to the environment and to what the animal is doing. It is possible that this mode of operation be a way of reconciling rest and alertness, allocating the attention according to the behavioural needs [61].

During desynchronised state, the stimulus evoked dynamics dominate the activity [96], and the representation capacity of the cortex is higher than during synchronised activity [111]. During synchronised state, the extended



stimuli are filtered out, and the slow oscillation dominates the dynamics, over the evoked activity [96]. However, punctual stimuli provoke down to up phase transitions and behavioural responses[30], [157]. One of the questions that arises- if we suppose that the sensory processing pipeline is serial - is: how can the upstream cortical circuits distinguish between an up-phase transition elicited by a sensory stimulus, from an internally generated one ? And more generally, in both states, how is the brain able to distinguish between the evoked and the spontaneous activity [157], knowing that the range of firing during the spontaneous activity is of the same magnitude as the range of evoked activity? We consider several hypotheses.

### 1.5.2.1 Noise subtraction hypothesis

One first hypothesis is that the brain subtracts out some kind of efferent copy, so we can't hear our own noise, as that it happens with tickling (we can't tickle ourselves). This hypothesis would fit well with some signal like the synchronous state, that spreads around all the cortex: it suffices that some downstream area process the difference between the feed-forward input from a lower area and from some neighbour cortical area that is not concerned with audition, but who is affected by the synchronous activity.

### 1.5.2.2 Different time scales over the cortical hierarchy filter out the non-perceptually related activity patterns

One important feature of the brain is that the information processing units, the neurons, receive currents, and emit spikes on a time scale of the order of milliseconds, whereas the time scales at which the behaviours happens is at least of the order of tenths of a second. One possibility to reconcile this apparent over-sampling of the world with the time scale on which the body acts on the world, is that somehow in order to form percepts, the incoming information is accumulated at different stages of the perceptual pyramid, and each intermediary unit of perception is only active when many of the units that project to it are active. In that way, as we go up the processing hierarchy, the time scale of the units increases also.

If we consider that the spontaneous activity as a sequence of incoherent neural patterns, we can think that only the stimuli, simply due to their temporal persistence in the environment, are going to repeatedly elicit sequences of identical patterns which are going to elicit higher order representations.

The randomly generated sequences of neural patterns, because of their temporal incoherence, are not going to pass through the different processing stages.

Murray et al. [106] show experimentally the existence of a hierarchical ordering in the time scale of the intrinsic fluctuations in spiking activity of spike trains recorded in monkeys engaged in cognitive tasks. The authors simply measured the width of the spike trains autocorrelation in different brain areas: in more sensory areas they found shorter time scales, whereas in pre-frontal areas, they found longer time-scales.

In their interpretation of this observation, the role of the time scales on the cortical hierarchy is similar to the one advanced here, even if they doesn't comment on the spontaneous activity: "shorter time-scales in sensory areas enable them to rapidly detect or faithfully track dynamic stimuli. By contrast, pre-frontal areas can utilize longer time-scales to integrate information and improve the signal-to-noise ratio in short-term memory or decision-making computations".

### 1.5.2.3 Signal and noise have different directions in neural space

Another possibility, is that the spontaneous and the evoked activities live in orthogonal subspaces. In their seminal paper, Kaufman and colleagues [78] proposed a simple but very elegant mechanism by which determined areas could isolate themselves from surrounding areas to do some local processing, despite the fact that they are physically connected to these areas and also communicate with them. During a reaching task, a monkey has to see a light on a screen and withhold from reaching it before a go cue appears. When the go cue is on, the monkey must touch the place where the light was, in order to receive a reward. Parallel recordings show that the premotor cortex is very active during the hold period. The question is: why this pre-motor activity doesn't trigger movement despite the fact that it projects to the motor cortex, whose activity correlates with the muscle activity ? They propose that as long as the activity in the pre-motor cortex lives in a null space, motor cortex doesn't see the difference. Lets take for example two neurons in pre-motor cortex which project to a neuron in motor cortex. The activity of this neuron innervates a muscle, with a strength proportional to its firing. As long as the sum of the activities of the pre-synaptic neurons is constant:  $p_1 + p_2 = (p_1 + \kappa) + (p_2 - \kappa) = C$ , the post-synaptic currents in motor

cortex are constant also. When the activity of both neurons increase, the direction that causes more effect on the post-synaptic current is the direction orthogonal to the null space, called the output potent space:  $(p_1 + \kappa) + (p_2 + \kappa) = C + 2\kappa$ . Then, we can have local processing - preparatory activity-happening in pre-motor that motor cortex doesn't see, and communication between the areas after the go cue.

Similarly it could be that the spontaneous activity lives on a subspace on a low area of the cortical hierarchy, but when a stimulus arrives, the circuit responds in such a way that the elicited activity has a certain direction that excites the next processing stage of the perceptual chain. Then, different processing areas might be isolated, and the spontaneous activity might serve as substrate for computations.

## 1.6 Possible roles of spontaneous activity

Even though there were early reports in the 1970's pointing at correlations between behaviour and the ongoing activity [79],[150] this activity was considered until the 90's like a noise [145], a fundamental limitation or a by-product of the system, which constrained the discrimination of sensory responses [157]. Some authors have argued against this noise hypothesis due to the important metabolic cost of spiking [5], suggesting that spontaneous activity might actually play a functional role in the brain. In what follows, we will review some of the advanced hypothesis.

### 1.6.1 Developmental role

Increasing evidence points at the fact that spontaneous activity might play a fundamental role in the development of neurons and in their connections.

In the mainstream theory of brain development, genetic programs organize the main projections, for example from the retina to the brain. Later, the visual experience leads to a refinement of the connections. However, more and more observations reveal that in fact genetic programs and neural activity interact at all phases of development and determine the composition and the organisation of neural circuits [16]. In early phases, when neurons are still not connected, spontaneous activity affects neuronal differentiation, establishment of neurotransmitter phenotype, and neuronal migration [140].

Later, spontaneous activity seem to have the role of providing an instructive signal for the establishment of the functional connections, for example in motor neuron path finding [58], or in the formation of sensory maps [77]. However as Blankenship and Feller point out [16], disambiguating the causality between the activity and development is not easy: "Insights into how spontaneous correlated activity influences the development of neural circuits will require manipulations that alter the pattern of activity rather than block it entirely".

### 1.6.2 Memory consolidation

Following severe seizures that had no treatment at the time, the patient H.M. went through a bilateral medial temporal lobe resection. Following the surgical operation, Scoville and Milner [135], performed several neuropsychological tests on H.M. and concluded that the hippocampus had a decisive role in the formation of new memories and in the retrieval of recent memories, but not in the retrieval of oldest memories, which remained intact. Influenced by these pioneering results, and by David Marr's [97] theory about memory, the actual theory about the role of the hippocampus is named the "two stage theory of memory". This hypothesis postulates that the memories are initially created and stored in the hippocampus. During sleep, these memories would be reactivated and transferred to the medial pre-frontal cortex **mPFC**, where they would be stored more permanently in a process called consolidation, that involves reorganization and strengthening of the cortical memory trace. In parallel, as the mPFC becomes more involved in recalling the memories, the hippocampus disengages progressively.

Experimental studies have brought evidence of an interaction between the hippocampus and the cortex that involves different types of waves. The first studies showed correlational evidence of a coupling between the transitions to the up phase of synchronous activity -that happen in cortex- and sharp wave bursts in the hippocampus [10],[138]. Bennett and colleagues [12] advanced the hypothesis that the condensed joint spiking in the up phases of the synchronous state might have a role magnifying post-synaptic responses and facilitating spike-timing dependent plasticity -**STDP**.

Very recently, the role of the coupling between hippocampus and cortex in memory consolidation was demonstrated causally [95]. Rats were first placed on an arena in which there were two objects. When sharp wave ripples

were detected in the hippocampus, the mPFC was stimulated electrically in closed loop. Finally, the rats were put back in the arena containing the objects, having previously displaced one of the two objects. To have a behavioural readout of the memory, they quantified the time that the rats spent close to the displaced object, which was interpreted as a readout of memory or novelty.

Indeed, it was the case that rats spent more time close to the displaced object when the stimulation protocol was precisely time locked, and not when the stimulation was applied with a random delay after the detection.

### 1.6.3 Bayesian inference

One very innovative hypothesis about what perception is and how it occurs, is that perception can be understood as a Bayesian inference process, i.e. a process that aims at deducing which objects we are experiencing in the external world, given noisy evidence and prior information about the context. As pointed out by Kok and de Lange [44], the motivation for developing such kind of models is that "perception is not solely determined by the input to our eyes, but it is strongly influenced by our expectations".

Lets imagine for example that  $x$  is a scalar variable that quantifies an unidimensional stimulus, for example a sound pressure wave of the voice of a person.  $x$  might come from two persons called 1 and 2, whose respective voices  $y_1$  or  $y_2$  may have different distribution of frequencies. Given that the environment is noisy, the perception that will try to infer is the joint distribution of the possible causes of  $x$ ,  $p(Y_1, Y_2|x)$ , through Bayes rule, and ultimately  $y_1$  or  $y_2$ , so that we know who is talking.

$$p(Y_1, Y_2|X) = \frac{p(X|Y_1, Y_2)p(Y_1, Y_2)}{Z}$$

$p(Y_1, Y_2)$  is the prior distribution of inputs, it describes the learnt regularities about the sensory environment, in this example the joint power spectrum of the two persons.  $p(X|Y_1, Y_2)$  is called the likelihood, it assumes that we have a model of the environment that allow us to evaluate how likely would the observation  $x$  be, if we knew the values of the two features  $y_1$  and  $y_2$ .  $Z = p(X)$  is a constant with respect to  $Y_1$  and  $Y_2$ . Finally  $p(Y_1, Y_2|X)$  is the posterior distribution, from which we deduce  $p(Y_1 = y_1, Y_2 = y_2|X = x)$ . Under this framework, perception results from the match between the prior

(from the regions where the prior  $p(Y_1, Y_2)$  is not nil), and the model of the world that we have, given the sensory evidence. If the prior is flat for example, the posterior will reflect simply our model of the world.

Lets imagine that the receptive field of one toy neuron is a frequency band of low pitch with a mean of 100 Hz, and a variance of 20 Hz. In time, the firing rate of this toy neuron fires at a mean of 100 Hz, and has a variance of 20 Hz: the instantaneous firing rate represents the probability of taking a certain value, as if this neuron was sampling from the distribution of frequency corresponding to the receptive field. Under this "sampling-based" hypothesis, if in the brain neural populations somehow use this representation, this framework [42] proposes an interpretation of what spontaneous activity might be: in the absence of sensory stimulation  $x$ , the posterior distribution is similar to the prior distribution, and the spontaneous activity reflects this prior. One of the predictions of this framework is that spontaneous activity is similar to the evoked activity, which has been observed many times [80], [43], [89] (we will comment more on this point). Another possible consequence is that on the motor side this spontaneous activity could have a role in preparing actions, so that animals are ready to respond faster. Overall, this bayesian sampling theory shows normatively how to combine the priors, the **top-down** expectations about the world with the sensory evidence about it, and precisely the spontaneous activity plays the role of reflecting those priors about the world.

A more mathematically involved theory called **predictive coding** [44], also based on Bayesian inference, conjectures that each region of the cortical hierarchy makes hypothesis about its incoming inputs. In this theory, each region generates a hypothesis about the incoming input, and feeds the error (the difference between the model and the input), to the upstream region, and feed-backs the more consistent hypothesis, to the downstream region. After several iterations, the predicted error is small in all regions and an hypothesis about the source that generated the sensory stimulus is retained. A whole literature testing some of the ideas of predictive coding exists. Let's only cite Sadaghiani and colleagues [130] which performed a detection task with humans on a magnetic resonance imager. The authors observed that consistently, in the trials in which the subjects reported having heard a dim sound, the spontaneous activity in auditory cortex that preceded the presentation of the sound was significantly higher in hit trials with respect to miss trials.

#### **1.6.4 Sampling information from the environment in a way that conciliates resting and alertness. Encoding of this information in packets that propagate around the brain**

According to Harris and Thiele [61]: "the brain continuously adapts its processing machinery to behavioural demands. To achieve this, it rapidly modulates the operating mode of cortical circuits, controlling the way that information is transformed and routed".

In 1975, Kemp and Kaada [79], using EEG, remarked a correlation between hippocampal theta rhythm (4-10 Hz) and attention: "The present observations of a striking relation between theta activity, scanning eye movements and intent staring behaviour indicate an involvement of theta activity in information sampling aspects of attentive behaviour". In 2009, Luczak [89] studying the temporal structure of cortical activity made observations that would lead to a similar hypothesis as Kemp and Kaada. They discovered that the sequential order of firing of cells -some cells fire earlier than others- during spontaneous activity in the inactive state is somewhat similar to the sequence during evoked activity. Depending on the stimulus, consistent small variations in the firing order are observed, but the coarse order of which neurons fire first and which neurons fire later is preserved.

In perspective of these findings, Luczak [90] proposed that the active state and the inactive state are simply related by the frequency of these packets - that trigger up states-. In the active state the frequency of these packets is high so that there are no down phases.

Given that -as we will also see later- the cortex becomes activated during attentive phases, Luczak, McNaughton and Harris [92] have advanced the hypothesis that the up and down-states correspond to discrete samplings of the thalamic input by the cortex, so that the drowsiness during awake correspond to moments in which the brain samples the environment, but with a lower frequency.

The interpretation proposed about these sequences of activity, or "packets", is that they are basic information units that are exchanged across the brain, like packets of bits that travel across the Internet. The Internet packets follow a protocol, a code of how to encode and to read these packets. The first bits are called the header, and convey the address they are going to, and the following bits convey the information. In the same way, Luczak

and colleagues postulate that the earliest cells in responding would convey, in their firing pattern, some information about the modality for example and then the later cells would encode the details of the stimuli. Because the up and down phases propagate all around the cortical mantle, but the stimulus-evoked up phases originate in different cortical areas, this mechanism might be a way of exchanging information across sensory modalities and across hierarchical areas of the cortex, in order to integrate effectively bottom up and top down signals .

## 1.7 Theoretical considerations about the mechanistic origin of the spontaneous activity

It is not currently known what is the origin of the spontaneous activity, however, due to the seminal work of Gerstein and Mandelbrot [49] and then of van Vreeswijk and Sompolinsky [149], there is a whole theoretical framework of the desynchronised state.

The majority of models of spontaneous activity are based on the premise that a neuron generates activity because it keeps receiving activity from its neighbours, and that as these neurons in turn generate activity, they project to other neurons which in turn generate activity and so on. Briefly said that the activity is **self-consistent**. We will examine next some examples of models of spontaneous activity both at the single level as at the network level.

### 1.7.1 Spontaneous firing at single cell level

We will start precisely by an example in which the self consistent hypothesis is not necessary, because in this theory cells generate spontaneously action potentials even in the absence of stimulation. It is known that voltage-gated ion channels open and close in a stochastic way, and that depending on the geometrical configuration of the cells, the spontaneous rate of action potentials can vary. For single compartment models of neurons, the neuron membrane potential is modelled as if it was homogeneously distributed on a sphere. When we consider large membrane areas, the effects of the stochastic openings and closings of the membrane potential are averaged out, and the somatic membrane potential is well described by the determin-



istic Hodgkin-Huxley equation [67]. According to biophysical simulations, for unmyelinated axons, the spontaneous action potential rate increases as the diameter of the axon thins. Also for both short and long axons the spontaneous action potential rate is high, whereas it is lowest for intermediate length axons [108].

In the sixties, Rodieck, Kiang and Gerstein [125] had the idea of studying the statistical properties of the spike trains associated with spontaneous activity of single neurons recorded from the cochlear nucleus of anesthetized cats. They observed different bursting patterns, from regularly spiking cells to bursting cells. When plotting the histograms of inter-spike intervals, they find unimodal and asymmetric distributions and also unimodal, symmetric and bell-shaped distributions, that when replotted on a logarithmic axis on the ordinates, resulted respectively in a rapid increase followed by a linear decay or by another bell shaped distribution. These observations suggest that the first kind of distribution might be generated from a Poisson process, whereas the second might be generated by a Gaussian time jitter process. Both Gaussian and Poisson processes have the interesting property that the successive interspike intervals are independent.

During a **Poisson process**, the probability of observing an event during the interval  $[t, t + dt[$  is  $\lambda \cdot dt$ , where  $\lambda$  is the rate of events occurrence. This means that the events are independent between themselves in time. It is analogous to a situation in which, at each time point, we threw a biased coin and we reported an event when it fell on heads. Precisely, a Poisson process has a **coefficient of variation**  $C_v = \frac{\sigma}{\mu}$  that is equal to 1. During a Gaussian time jitter interval process, the inter-spike interval has a defined mean and is normally distributed, so that the spike trains show some periodicity.

Gerstein and Mandelbrot [49] had then the idea of modelling the behaviour of previously observed cells as a **random walk model** with drift. In this model **epsp**'s and **ipsp**'s -excitatory (resp. inhibitory) post-synaptic potentials- arriving randomly at similar rates break the history dependence between spikes, because the random arrival of positive and negative post synaptic potentials provokes a forgetting effect on the membrane potential with regard to its initial condition. This model did reproduce the Poissonian inter-spike interval histograms of two cells as well as the ISI of the cell that had a more Gaussian distribution [125].

Analysing data from monkey visual cortex ( $V_1$  and  $MT$ ), Softky and Koch [139], also measured that the coefficient of variation in cortical cells

with high firing was close to one  $C_v \approx 1$ . Misunderstanding the most important prediction that Gerstein and Mandelbrot had formulated 30 years before, which was to use both excitatory and inhibitory random input currents, Softy and Koch fed different kinds of **integrate-and-fire** models with only excitatory post-synaptic potentials -**epsp**'s of identical amplitude generated at random times from a Poisson distribution, and computed analytically or simulated the inter-spike interval distributions. Softky and Koch choose to use randomly generated inputs because when neurons are injected with a constant current, neurons spike in a regular fashion [94], therefore the spiking irregularity in-vivo must be caused by fluctuations in synaptic inputs.

Considering the central limit theorem, they understood that in the simulations, when adding an elevated number of randomly generated epsp's, it was expected to observe a highly stereotyped inter-spike interval, and then a very low  $C_v$ , completely at the opposite of what they observed experimentally. They concluded from this, that the integrate-and-fire models -which suppose that the fundamental operation that neurons do is integration- is an abstraction that is too simple and then added active dendrites to the biophysical models, which allowed them to re-obtain a coefficient of variation around 1.

We now think that the prediction of Gerstein and Mandelbrot regarding the role of the interplay between the excitation and the inhibition in the generation of random spike sequences during spontaneous activity, is crucial. Also, the integrate and fire model is thought to be a very good model of spike generation [18].

Finally, lets mention some models that explain the existence of bi-stability in single cells firing, i.e., the possibility of one cell to fire, for a same input, with two different firing rates, depending on the history of the inputs [54].

### 1.7.2 Spontaneous activity as a neural network property

The previous models of spontaneous activity were based on single cell properties. In the 60's, when Gerstein and Mandelbrot [49] formulated their drift diffusion model, they were already aware that some physiological variables ("temperature, anesthesia, bio-chemical manipulation, sleep or wakefulness, "states" of alertness") could have an impact on the spontaneous discharges.

However, as we said in the introduction, the brain state is a macroscopic quantity, that depends on the way cells fire with respect to each other. We will go over the models of synchronous states keeping in mind, as we said above, that these models receive external noisy activity from the outside.

For didactical purposes, we will present first the models related to the asynchronous state, which contrary to the synchronous state, took a long time to understand, and that led to the concept of balanced state by van Vreeswijk and Sompolinsky [149], and which was extended by Renart and colleagues [123], to account for the near zero mean pairwise correlations in the asynchronous state.

Amit and Brunel [4] introduced the idea that **persistent activity** (i.e. activity following the presentation of a stimulus, during a memory task), would happen at the network level as small deviations around a fixed point or **attractor**. In the context of the memory, we understand that these attractors represented the memories, in the form of patterns of activity. A neural network is a complex system described by coupled differential equations, and the persistent activity would result from the small perturbations around the attractors. This is indeed the way we think about the active state, like small perturbations around a fixed point. Amit and Brunel found that a network composed only of excitatory neurons randomly connected and receiving stochastic inputs from the outside is not stable, but when fast inhibition is added, the network becomes stable, and shows a physiologically plausible spontaneous rate.

### 1.7.2.1 The balanced state

The observation of Softy and Koch [139] that the irregular firing of cortical neurons required an explanation triggered a lot of attention, which resulted in a series of theoretical studies proposing different alternatives. Forty years after Gerstein and Mandelbrot [49]- Shadlen and Newsome proposed that if the excitatory and inhibitory presynaptic inputs were balanced and effectively cancelled each other on average, the resulting fluctuations would trigger spikes that looked temporally irregular [136], [137]. Related ideas were explored by Bell et al, Tsodyks and Sejnowsky [148] and Troyer and Miller [146]. However, none of this studies provided a generic explanation of how pre-synaptic excitation and inhibition could *become* balanced in a robust way which was tolerant to differences in synaptic strength or other

parameters of the network connectivity. In their description of the **balanced state**, van Vreeswijk and Sompolinsky [149] provided such an explanation, which has become a cornerstone in our understanding of cortical dynamics and of tonic spontaneous activity.

The network is composed by two groups of excitatory and inhibitory neurons and one group of external neurons generating only excitatory input. Apart from that, the neurons are connected with a certain probability and random weights. The network is said to be randomly connected.

This model postulates that the origin of the spiking irregularity comes from a high level of pre-synaptic excitation and inhibition which are both large compared to the neuronal threshold and which cancel each other out, resulting in a sub-threshold membrane potential, whose random temporal fluctuations are of the same order as the mean. Because of this cancelling, which happens dynamically (i.e independent of the details of the network), a small increase in excitation is compensated by a rapid increase of the firing of inhibitory cells, keeping constant the overall firing rate.

We will briefly replicate the **balance condition** without providing all the context that leads to it. When connected, the connection strength between two neurons is equal to  $\frac{J_{ZW}}{\sqrt{N}}$ , where  $ZW$  designates that this connection goes from the pre-synaptic  $W$  to the post-synaptic  $Z$ , and  $Z, W = E, I, X$  (excitatory, inhibitory, external).  $N$  is the size of the network. The scaling of all the connections in the network is of order  $O(\frac{1}{\sqrt{N}})$ : this means that they converge to zero as  $\frac{1}{\sqrt{N}}$ .  $f$  is the fraction of excitatory cells among the population. We call  $\mu$  the mean post synaptic current and  $\langle r \rangle$  the mean rate. The mean post synaptic current for the excitatory and for the inhibitory neurons can be expressed as:

$$\begin{cases} \mu_E = \sum_{j=1}^{f.N} \frac{J_{EE}}{\sqrt{N}} \cdot \langle r_E \rangle + \sum_{j=f.N+1}^N \frac{J_{EI}}{\sqrt{N}} \cdot \langle r_I \rangle + \sum_{j=1}^N \frac{J_{Ex}}{\sqrt{N}} \cdot \langle r_X \rangle \\ \mu_I = \sum_{j=1}^{f.N} \frac{J_{IE}}{\sqrt{N}} \cdot \langle r_E \rangle + \sum_{j=f.N+1}^N \frac{J_{II}}{\sqrt{N}} \cdot \langle r_I \rangle + \sum_{j=1}^N \frac{J_{Ix}}{\sqrt{N}} \cdot \langle r_X \rangle \end{cases}$$

Developing :

$$\begin{cases} \mu_E = \sqrt{N}(J_{EE}f \cdot \langle r_e \rangle + (1-f) \cdot J_{EI} \cdot \langle r_i \rangle + J_{Ex} \cdot \langle r_x \rangle) = \sqrt{N} \cdot C \\ \mu_I = \sqrt{N}(J_{IE}f \cdot \langle r_e \rangle + (1-f) \cdot J_{II} \cdot \langle r_i \rangle + J_{Ix} \cdot \langle r_x \rangle) = \sqrt{N} \cdot D \end{cases}$$

The main insight is that, in the large  $N$  limit, to be in the balanced state

$C, D$  have to be  $O(\frac{1}{\sqrt{N}})$ . Setting  $C$  and  $D$  equal to zero and solving the equation, we obtain:

$$\begin{pmatrix} \langle r_e \rangle \\ \langle r_i \rangle \end{pmatrix} = - \begin{pmatrix} f J_{EE} & (1-f) J_{EI} \\ f J_{IE} & (1-f) J_{II} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \langle J_{Ex} \rangle \\ \langle J_{Ix} \rangle \end{pmatrix} \cdot \langle r_x \rangle$$

As we want the firing rates to be positive we have both conditions : numerator and denominator must be of the same sign. This leads to two possible conditions :

$\frac{J_{EX}}{J_{IX}} > \frac{|J_{EI}|}{|J_{II}|} > \frac{J_{EE}}{J_{IE}}$  or  $\frac{J_{EX}}{J_{IX}} < \frac{|J_{EI}|}{|J_{II}|} < \frac{J_{EE}}{J_{IE}}$ . As long as the mean connectivity values obey these conditions, the excitatory and the inhibitory rates will reflect the input and will cancel each other, in such a way that the expressions  $C$  and  $D$  above will be order  $O(\frac{1}{\sqrt{N}})$ , and the network will be in the balanced state.

In addition of showing approximate Poissonian statistics of the inter-spike interval at single cell level, the balanced state makes an important additional prediction of the distribution of rates in cortical circuits. The strong neuron-to-neuron variability in mean synaptic input, plus the expansive non-linearity of single cell f-I curve leads to a wide distribution of firing rates which shows a characteristic skewed shape with a long tail. Another strong prediction of this model -although hard to prove experimentally- is the **chaotic** character of the cortical networks in the balanced state. This means that small variations in the state of the network -like flipping the state of one neuron from off to on- induce with time, exponentially large differences of the state of the network, i.e, the network pattern of activity will be completely different after a finite amount of time.

### 1.7.2.2 Strongly connected densely coupled networks generate an asynchronous state

The fourth and the second-third cortical layers receives direct feed-forward input from the thalamus. Inside each layer, neurons are recurrently connected, and layers are connected between themselves: for example layer IV projects in a feed forward way to layer 2/3. As we will see in the chapter "Cortical neurons integrate common input from sensory thalamus", pairs of cells in the fourth layer have a very high probability of connection up to 50

%, and thalamic axons coming from the thalamus tend to target preferentially pairs of cells that are connected. The cortex anatomy seems to enforce positive correlations between pairs of inputs.

On another side, given that the neural responses are noisy from trial to trial, a good way of to get rid of this noisiness, would be to use redundancy, i.e. encoding in parallel the response to a given stimulus in different cells of a population, and then averaging all those responses in a downstream structure [136]. Depending of how this encoding is done, this might be or not a good way of shedding the single cell trial to trial variability.

If, when one stimulus is presented repeatedly, cells respond from trial to trial in a different way, but correlated among them, this is precisely a bad thing because on a single trial the noise will not be averaged out on a downstream cell. The variance of the population rate give us an estimation of the error that a downstream structure would make estimating this population rate, by pooling neural responses from an entire population. Lets call  $V$  the variance of the population rate of  $N$  neurons:  $V = \text{var}(\sum_{i=1}^N \frac{n_i(t)}{N})$ . Lets assume for simplicity that all the neurons in the population have the same variance  $v$ . Lets finally call  $\bar{r}$  the mean covariance/correlation between all pairs in the population. If we develop the definition of the population rate variance, we obtain [168]:

$$\frac{V}{v} = \frac{1}{N} + \bar{r}$$

If the activity of neurons is not independent,  $\bar{r}$  is finite and independent of  $N$ : the network is said to be in a synchronous state. If neurons are independent,  $\bar{r} = 0$ , the bigger the population size, the smaller the error in decoding a stimulus might be. It might also be that  $\bar{r}$  scales as  $\frac{1}{N}$ , and then even if there are finite correlations the error might be done arbitrarily small. In both previous cases, the network is said to be in an asynchronous state [50], [123].

Renart and colleagues [123], showed theoretically that it was possible to build a balanced recurrent network that received large amounts of shared inputs, and that still attained a near zero mean pairwise correlation, that scaled like  $\frac{1}{N}$ . When excitatory inputs are either shared or correlated, they generate positive correlations, and the same applies to inhibitory inputs. But if excitatory and inhibitory presynaptic inputs are positively correlated, they generate negative correlations of this nature which cancel precisely all sources of positive correlations induced by the connectivity and the dynam-

ics. The balance condition [149], ensures that inhibition tracks excitation, and therefore that the inhibitory and the excitatory input currents tend to be positively correlated [123], decorrelating the network and maintaining it in asynchronous state. The theoretical argument to prove this tracking at the level of the currents, follows a similar logic than the balance condition for the firing rates, presented previously. The difference between the network of Vreeswijk and Sompolinsky and the network of Renart and colleagues, is that at the level of the connectivity, the first one is called sparse (the connection probability between two cells scales as  $\frac{1}{N}$ ), whereas in the Renart et al. network, the connectivity is said to be dense, i.e. independent of the network size.

Tetzlaff and colleagues, show that one can also obtain a mean correlation that is of the order  $O(\frac{1}{N})$  in inhibitory-only networks [144].

### 1.7.2.3 Influence of the negative feedback on different dynamical regimes

The asynchronous state is such that the inhibition can track the excitation, so any small deviation in the firing is compensated almost immediately by the inhibitory feedback, in such a way that the global firing remains constant. Asynchronous state has a **fast negative feedback**.

The key ingredient in generating up and down phases during synchronous state seems to be the **adaptation** [23], also known as **spike-frequency adaptation**. Injecting a square pulse of current leads first in some cells to an initial burst of action potentials followed by a period of silence, in which the membrane potential is hyperpolarised, and then again to a new burst of spikes. Many possible biophysical mechanisms can lead to adaptation, but all of them implement a form of **slow negative feedback** of cell excitability. A simple analogy of slow feedback is an electric shower head in which when we open it too much, the water comes cold and then we rapidly close the knob; after some seconds the water comes too hot, and then we open back the knob: the delay in the feedback leads to an instability. There are many computational models of synchronised activity which use adaptation and which are able to reproduce many of the features of the data [27],[84],[85],[32],[101]. Some biophysical mechanisms that lead to adaptation are: inactivation of the sodium channels -which are responsible for the depolarizing currents-, activation of shunting currents ( $K^+$  channels, modu-

lated by acetylcholine) and hyperpolarizing potentials due to an increase in intracellular  $Ca^{++}$  concentration after a burst of activity [55].

### 1.7.3 Possible sources of correlations

When measuring the spiking of a pair of neurons, we should be aware that there are many biophysical factors that can contribute to their correlation. The arrangements of possible recurrent -or indirect- connectivities that can lead to same co-fluctuation strength are infinite. Lets cite the two most commonly discussed:

- Ko et al. [82] measured in-vivo activity, and identified the neurons that were strongly correlated, and then killed the mice and probed the direct connections between these neurons with patch-clamp, which resulted to be very strong. Very strong coupling between units might lead to positive correlations.

- Axons from an excitatory or an inhibitory neuron that project to a pair of cells, also known as common input, has repeatedly been considered a source of positive correlations [7], [137].

Conversely, Renart et al [123] showed that if the presynaptic inputs to a couple of cells are excitatory and inhibitory and positively correlated, then this causes decorrelation in the firing of both cells.

Another possible source of correlations is that the respective synaptic inputs to each of the measured cells are themselves correlated [159]. We can mention some other concrete elements of cellular and circuit structure that do affect the correlations: the neuronal thresholds distributions of the pre-synaptic inputs, the non-linearities that transfer input currents to spike outputs, the variability in the amount of vesicle release, the noise on the membrane potential. In the chapter 4, I did a detailed study of the effects of many of these biophysical parameters in the correlation between the membrane potential of two neurons, and analysed an experimental situation in which we can deduce the proportion of shared axons, independently of the other parameters.

As pointed by Doiron et al [37], "different mechanisms could explain the correlations observed under a single state. However, only some of these mechanisms will be consistent with observations from multiple states". By state, Doiron means the context under which neural activity is recorded, not necessarily cortical state. In [37], Doiron et al. develop a mathematical



framework that accounts and distinguishes for different biophysical factors that impact the correlations, even if the list is not exhaustive.

### 1.7.4 Amplification in cortical circuits: spontaneous activity reflects idiosyncratic features of the network connectivity

Several groups have reported, a similarity between the patterns of activity observed during the spontaneous and during the evoked activity [89],[147],[13], [43]. One hypothesis that might explain these diverse observations is that, due to learning, the connectivity in the cortical circuits may have adjusted to better process the statistics of the environment [13], and to **selectively amplify** certain stimulus with respect to others [105], [46]. Berkes and colleagues [13] report an increasing match between evoked and spontaneous activity during development.

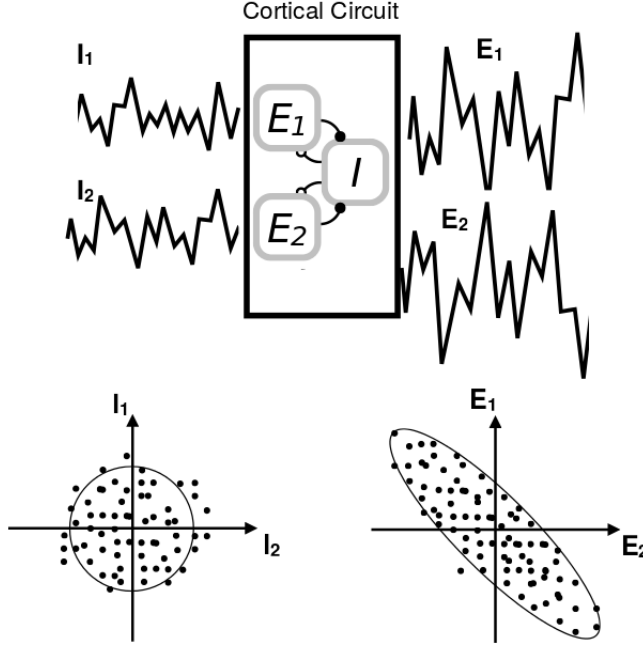
The patterns of activity that occur during spontaneous activity may be a side effect of the constraints that the connectivity impose on the expressed activity patterns [46]. To make this point clear, imagine the hypothetical situation in which a cortical circuit receives uncorrelated activity from its neighbour circuits (figure 1.5). Given the connectivity this circuit has, only certain activity patterns will be expressed. In the figure 1.5, two populations of excitatory cells interact through an inhibitory population, constraining the two populations of excitatory cells to have uncorrelated activity.

Given this particular connectivity, let's imagine  $\vec{v} = \begin{pmatrix} 1 & -1 \end{pmatrix}^t$  a vector giving the firing rate of each of the two excitatory groups with respect to their mean rate. A useful operational definition of the **amplification** of the pattern  $\vec{v}$  is that the population activity  $\vec{p}_{(t)}$  is dominated by a multiplicative temporal modulation  $\alpha_{(t)}$  of the pattern  $\vec{v}$ :

$$\vec{p}_{(t)} \approx \alpha_{(t)} \cdot \vec{v}$$

In the previous example we only choose one amplified pattern  $\vec{v}$ , but in the general case there could have been more amplified patterns  $\vec{w}$ ,  $\vec{x}$  ...

The definition of amplification we just gave embrace also the common knowledge usage of the word amplification in which a unidimensional signal  $s_{(t)}$  is multiplied by a scalar  $K$ ,  $\alpha_{(t)} = s_{(t)} \cdot K$ , but it expands it to the context of a multivariate system in which some joint patterns of activity



*Figure 1.5: Amplification in cortical circuits. To model spontaneous activity, we suppose that a local cortical circuits receives uncorrelated noise from the neighbouring circuits. Then, the most amplified patterns during spontaneous activity will result from the interaction of the noise with the recurrent connectivity. Indirectly, looking at the spontaneous activity tells us something about the circuit connectivity.*

$\vec{v}, \vec{w} \dots$  are multiplicatively modulated over time. In the figure 1.5, we see how the activity along the direction  $\begin{pmatrix} 1 & 1 \end{pmatrix}^t$  is filtered out, but also how the magnitude of the fluctuations along the direction  $\begin{pmatrix} 1 & -1 \end{pmatrix}^t$  is enhanced.

## 1.8 Aim of the thesis

In this dissertation we examine the temporal structure of spontaneous activity during states of cortical activation. Whereas a long series of studies have reported temporal structure during spontaneous activity, the large majority of the examples refer to activity during synchronised/inactive states. We have found that, whereas previous work had emphasized the lack of temporal structure during cortical activation, spontaneous activity during the active state shows a large degree of temporal coordination. We show that indeed there is a lack of temporal coordination *on average*, but that going beyond the coordination of individual pairs reveals a signature of competitive amplification of spontaneous activity fluctuations. We characterize extensively these competitive dynamics.

In the third chapter we study what could be the mechanistic basis for this type of amplification. We consider the necessity of structured connectivity and propose a circuit connectivity motif that reproduces many of the observed dynamical properties of competitive activity, using linear dynamical systems theory. Furthermore, we will show that the data during active state shows signatures of a particular kind of amplification called non-normal amplification.

In the fourth chapter we will discuss results from a study of functional anatomy which reveals patterns of common input across thalamo-cortical inputs to different layers. In particular, we will show that in the projection from the thalamus to the visual cortex, there is a spatial selectivity of those shared inputs. The long range projections that target layer IV of the cortex, target preferentially cells that are connected between themselves. We will go over the results of the study and then focus on modelling the impact of the proportion of shared inputs on the correlation between pairs of cells.

In the fifth chapter, we will compare statistical methods that try to determine the number of principal components in a multivariate data set, which is equivalent, in the context of amplification, to determining how many amplified patterns of activity  $p$  dominate the population dynamics :

$$\vec{p}(t) \approx \sum_{i=1}^p \alpha_{(t)}^i \cdot \vec{v}_i$$

In the last chapter we will summarize, discuss the results, and expand the scientific challenges.



# Chapter 2

## Temporal competitive structure during the desynchronised state

**HIGHLIGHTS**

- We find that during the desynchronised state, the activity is close to one dimensional, bringing to light two populations of neurons with graded anti-correlated activity. These dynamics are called competitive.
- We characterise the competition at the level of the single cells and also at the temporal and spatial level of the competitive dynamics.

### Contents

---

<b>2.1</b>	<b>Introduction: temporal structure during spontaneous activity . . . . .</b>	<b>38</b>
<b>2.2</b>	<b>Results: competitive activity revealed during the desynchronised state . . . . .</b>	<b>40</b>
2.2.1	Known features of the desynchronised state . . .	40
2.2.2	Raw phenomenon: competitive dynamics . . . .	42

2.2.3	How stable are the PC directions across the recording ? . . . . .	46
2.2.4	Do all neurons participate in the competition ? . . . . .	48
2.2.5	Role of global fluctuations in competitive activity . . . . .	50
2.2.6	Single cell characteristics . . . . .	56
2.2.7	Temporal characterisation of the competition . . . . .	60
2.2.8	Spatial characterisation of the competition . . . . .	64
<b>2.3</b>	<b>Discussion . . . . .</b>	<b>70</b>
2.3.1	Summary . . . . .	70
2.3.2	Temporal invariance of the competition . . . . .	70
2.3.3	Physiology, connectivity and competition . . . . .	71
2.3.4	Functional significance of competitive dynamics during spontaneous activity . . . . .	72
2.3.5	Low dimensional dynamics in cortex . . . . .	78
<b>2.4</b>	<b>Methods . . . . .</b>	<b>79</b>
2.4.1	Recordings and experimental procedures . . . . .	79
2.4.2	Preprocessing the spike times into spike counts . . . . .	80
2.4.3	Coherence methods for determining the relative position of the shanks . . . . .	81
2.4.4	Methods on Spectral Analysis . . . . .	84
2.4.5	Non-parametric statistical tests . . . . .	85
2.4.6	Necessity and sufficiency of a given subspace of activity . . . . .	88

---

## 2.1 Introduction: temporal structure during spontaneous activity

As we saw in the introductory chapter, spontaneous activity might be implicated with several fundamental processes like development, memory consolidation, sensory processing and behavioural modulation. It is therefore necessary to understand it better to have clues about its function. One

of the possible ways to study spontaneous activity is to look for **temporal structure**, i.e. for patterned organisation of the population activity in time.

For instance, Plenz and colleagues have worked on measuring a phenomenon called **neuronal avalanches** in superficial cortical layers during spontaneous activity. Guided by branching process theory, they measured the number of activated neurons between two silences, and observed that the ratio of descendants to ancestors is equal to one, which according to this theory, puts the cortex in a state called critical that confers special properties to the neuronal networks like optimal information transmission, capacity storage and computational power [117].

Since the 90's, Abeles and colleagues [120], have been studying precise firing sequences known as **synfire chains** in the deep layers of the cortex. The occurrence frequency of these events has been shown to correlate with the level of synchronisation. These sequences of activities have also been observed at the level of the membrane potentials in slices of mouse  $V_1$  and in cat  $V_1$  [74]. Their interpretation of this phenomenon is that this temporal structure is ideal for information transmission without loss, given that the cortex is composed of weak, stochastic, and depressing synapses. Moreover, it seems that these chains concatenate themselves in different orders to form **cortical songs**, which appear in a compressed format, as if the circuits were replaying previously learnt sequences [74]. A latter study [102] claims however that these temporal sequences are generated by chance, as a result of the constraints imposed by the dynamics of subthreshold spontaneous activity.

Both during inactive state as well as in the active state, Luczak and colleagues [89], [14], find sequences of firing that are similar to the evoked activity that follows the presentation of a stimulus. Bermudez Contreras et al. [14] show that during induced desynchronised state, the similarity between the spontaneous and the evoked activity patterns increases significantly after the stimulation with respect to before the simulation.

Using slices, Cossart et al. showed that the order in which neurons fire during the transitions from the down to the up phase of the inactive state follows a sequential structure [28]. Luczak et al [91] replicated this result in awake and anaesthetised rats, and showed that the temporal precision of the spiking order decays as the up phase progresses.

Okun and colleagues [109] understood in 2015 that when the population rate variability is higher, -i.e when the activity is more synchronised-

neurons fire in a consistent manner with respect to the population rate  $P(t) = \sum_{i=1}^N n_i(t)$ . This relationship happens at the level of the **covariance**. The covariance measures on average how much two random variables tend to vary together around their respective mean:  $cov(P(t), n_j(t)) = E((P(t) - \mu_P) \cdot (n_j(t) - \mu_{n_j}))$ . At the level of the populations there is a continuous gradation in the way neurons are correlated with the population rate  $P$ , some are very correlated while some others are less correlated (and have a bigger or smaller **population coupling**). Okun et al show that if one knows the population coupling (a scalar) from each single neuron, one can provide a good approximation of the covariance matrix of the whole population. This implies that the temporal structure of spontaneous activity in this conditions is low (almost one) dimensional. Also, the population rate is a good predictor of the sensory responsiveness, which is further explained by the finding that cells with strong population coupling receive more synaptic inputs from their neighbours.

## 2.2 Results: competitive activity revealed during the desynchronised state

### 2.2.1 Known features of the desynchronised state

We will first recall the main features of the desynchronised state using a representative experiment that we will use consistently in this chapter. As we see in figure 2.1, the desynchronised state is characterised by the following features: a fairly constant mean population rate (A), neurons fire tonically over time (B), the histogram of mean firing rates is uni-modal and skewed with a heavy tail (D). On average, the correlation between the neurons is close to zero (C), but slightly positive [123]. Finding near zero pairwise correlations in the cortex, generated much controversy in the 2000's, because there was on one side increasing anatomical evidence showing that nearby cortical neurons receive a substantial amount of common input, and at the same time there was a strong belief that common input lead necessarily to positive correlations between neurons [39]. The near zero average of pairwise correlations means that on average, there are as many pairs of cells that are positively correlated as pairs that are negatively correlated. This fact is consistent either with the scenario that neurons spike more or less



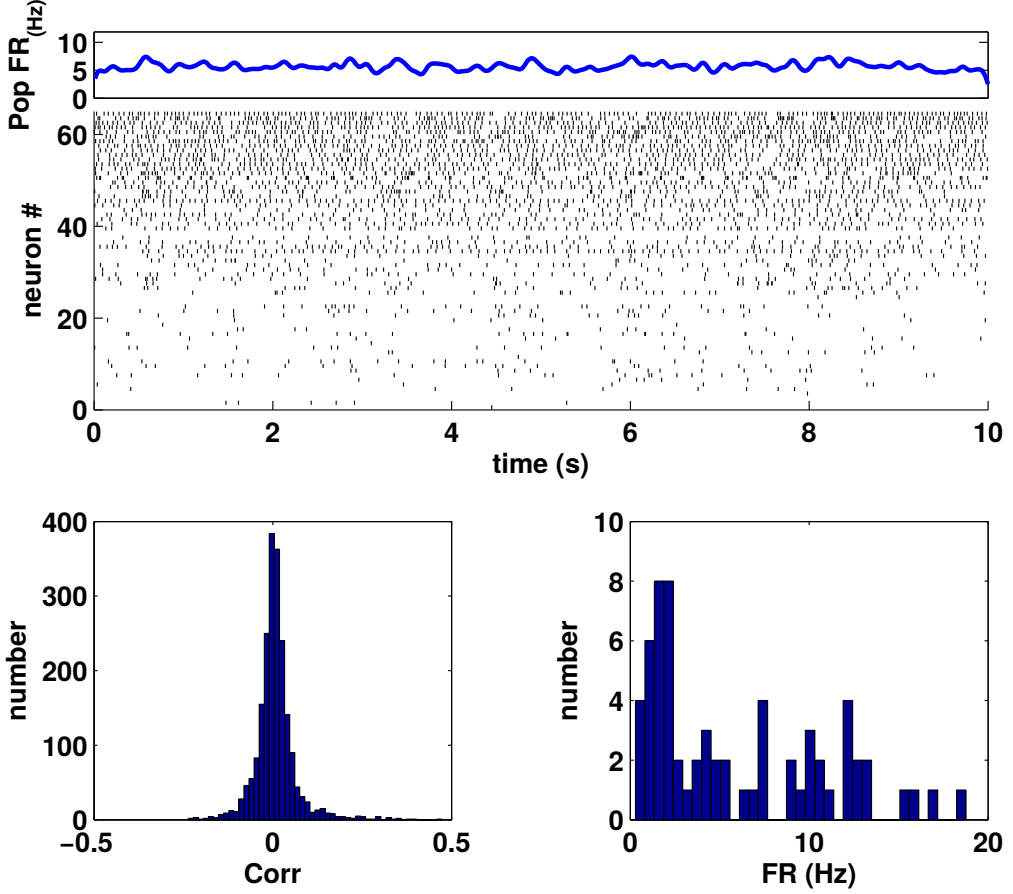


Figure 2.1: Desynchronised state in layer V of the cortex (S1). A: population rate. B: raster plot. Cells are sorted by firing rate. C: histogram of pairwise correlations. D: Histogram of firing rates. Experiment 065a. Details: A and C, Gaussian kernel ( $\sigma_1 = 0.1s$ ). D-H:  $t_c = 0.1s$ ,  $N_{jitt} = 10$  (see methods 2.4.2).

independently with finite correlations, which has been the current opinion, or as we will see, that each half of the population is positively correlated with itself and negatively correlated with the other half.

### 2.2.2 Raw phenomenon: competitive dynamics

In figure 2.2, we plotted the same 10 seconds presented in figure 2.1 but ordering the neurons differently. As we said in the previous paragraph, until now the second order analysis of spiking activity during the desynchronised state was limited to their distribution, ignoring the spatial correlations. We go beyond this approach by considering the correlation matrix and applying principal component analysis -**PCA**- to it.

When PCA is applied to the correlation matrix (figure 2.2 F), we see that there is a direction of the activity in the neural space that stands out with respect to the other ones. Reordering the correlation matrix and the raster plot (figure 2.2 C,I) with the same order as the sorted PC1 components, or **loadings** (figure 2.2 B), reveals structure in the population activity. On each diagonal of the correlation matrix neurons are positively correlated with their neighbours and negatively correlated with the neurons of the opposite side.

Because the dynamics of the population is dominated by anti-correlation, we will refer to this type of structure as "competitive", with excess activity in neurons with large positive PC1 loadings being associated with excess inhibition of neurons with large PC1 loadings -and vice-versa. As the variance associated to this competitive dimension is clearly standing out from the rest (figure 2.2 G), we say that the dynamics display **competitive amplification**.

In figure 2.2 D we see the instantaneous firing rate of the group of cells having a positive loading and the instantaneous firing rate of the group of cells having a negative loading: we see that the firings are anti-correlated. However, is the competition taking place between two separable groups? When we look at the histogram of the PC1 loadings, it appears unimodal (figure 2.2 E). Thus, rather than two groups, it seems like a continuum in such a way that neurons at either end of the continuum are strongly correlated and weakly correlated with neurons in the middle of the continuum.

Because the correlation matrix is symmetric, it can always be written as a sum of the external products of its eigenvectors (PC's):

$$C = \sum_{i=1}^N \lambda_i PC_i PC_i^t$$

$\lambda$  is the projected variance associated to each PC. In figure 2.2 J, we plot-

ted the rank one approximation of the correlation matrix:  $C \approx \lambda_1 PC_1 PC_1^t$ .

We can see that it is quite close from the original ordered correlation matrix, and therefore that the product between the loadings of two cells seems to determine the correlations to a very good extent. We then understand why the correlation matrix has such a structure, the  $PC_1$  being a graded vector approximately symmetric around zero (figure 2.2 B), the outer product of this vector with itself determines a matrix whose first and third quadrants are positive and whose second and fourth quadrants are negative. In the figure 2.3, we plotted all the eigenspectra, all the correlation matrices and all the PC1, for all the experiments. Each experiment corresponds to a different rat.

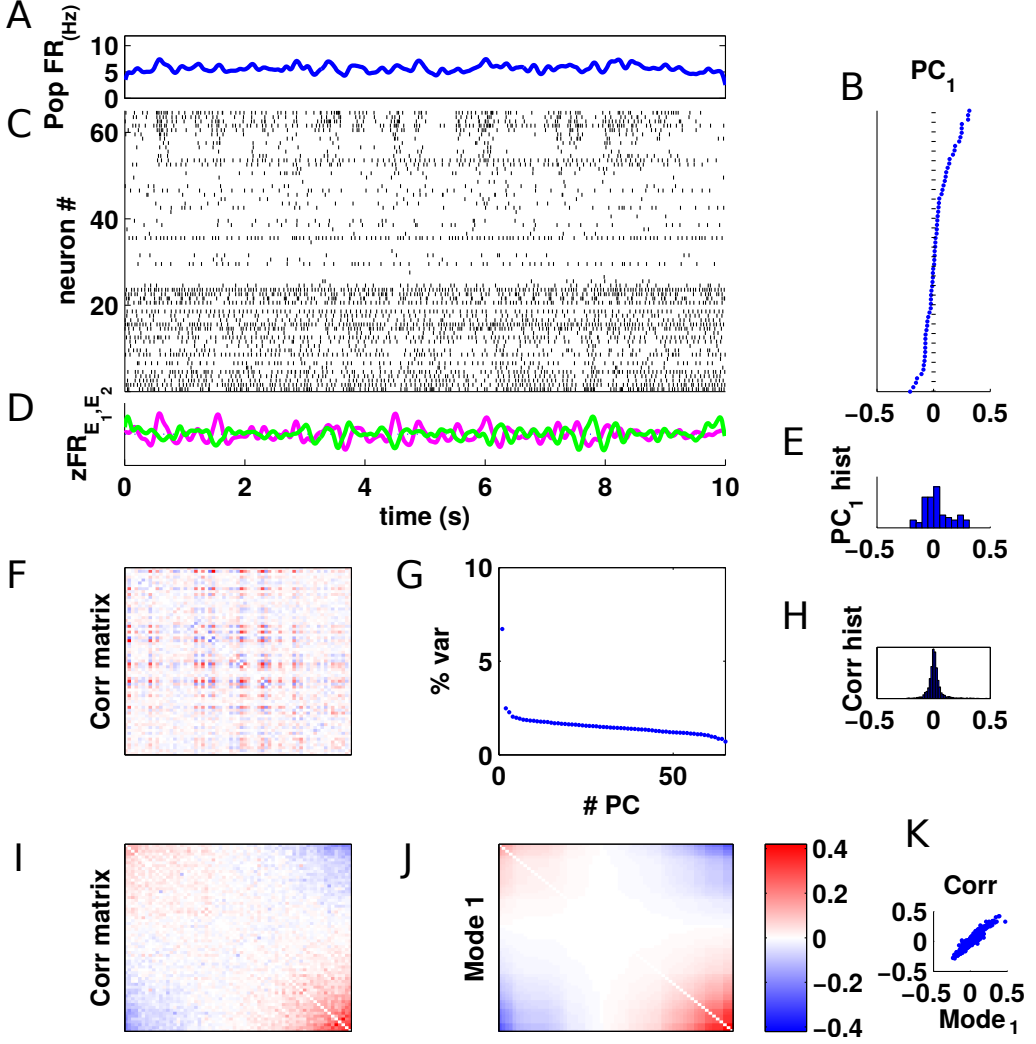


Figure 2.2: A: population rate. B: PC<sub>1</sub> components. C: raster plot, same experiment and same time as in figure 2.1, sorted using the order given by the PC<sub>1</sub> coefficients. D: z-scored population rate of each of the two populations. E: histogram of PC<sub>1</sub> loadings. F: correlation matrix. G: eigenspectrum. H: histogram of correlations. I: sorted correlation matrix using the order of the PC<sub>1</sub> loadings. J: first mode of the correlation matrix. K: correlations versus mode 1 coefficients. Experiment 065a. Details: A: Gaussian kernel  $\sigma = 0.1s$ , D: Mexican hat kernel ( $\sigma_1 = 0.1s, \sigma_2 = 4\sigma_1$ ). B, E, F, G, H, I, J:  $t_c = 0.1s, N_{jitt} = 10$  (see methods 2.4.2).

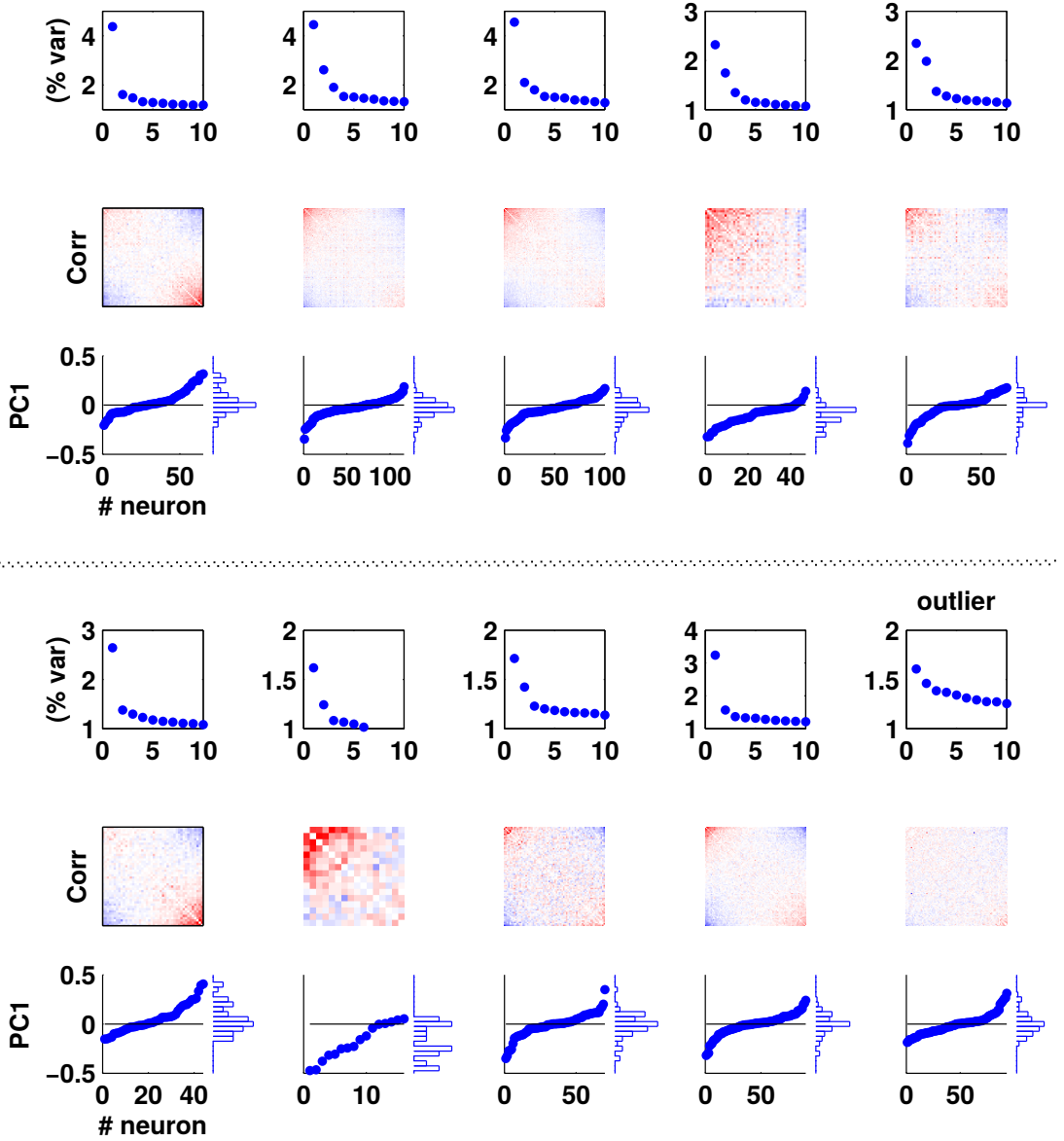


Figure 2.3: Competitive activity across experiments ( $n = 10$ ). First row: percentage of explained variance for different experiments (rows). Second row: correlation matrices ordered by the PC1 loadings. Third row: PC1 loadings and histogram of loadings. Parameters:  $t_c = 0.1s$  and  $N_{jitt} = 10$ .

### 2.2.3 How stable are the PC directions across the recording ?

In order to make sure that the structure shown in the previous section is not a statistical fluctuation, we can split the recordings in halves and assess how does the direction of the principal components stay stable along the recording, by computing the alignment of the  $n^{\text{th}}$  principal component computed on the first half of the recording with the  $n^{\text{th}}$  principal component computed on the second half of the recording. The alignment is then a scalar quantity between 0 and 1 that quantifies the persistence of a direction of activity along the recording:  $PC_n$ ,  $n \in [1, N]$

$$|(PC_n^{(1)})' \cdot PC_n^{(2)}|$$

For each experiment, we have then only one overlap that we plot in figure 2.5, for the first ten principal components.

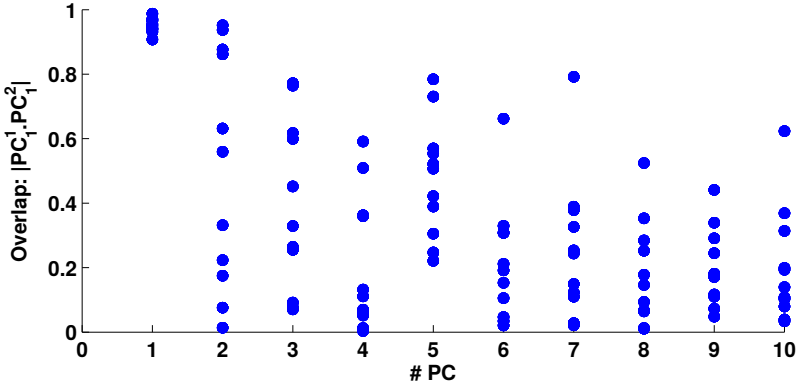


Figure 2.4: A: Overlap of the  $n^{\text{th}}$  principal component computed in the first half and the  $n^{\text{th}}$  PC computed in the second half of the recording, for each experiment. Parameters:  $t_c = 0.1s$  and  $N_{jitt} = 10$ .

This method allows us to see that at the level of the recording there was not some kind of drift between the begin and the end that would change the direction of the PC1. However, this method is very strict, in the sense that for example  $PC_3$  in the first half of the recording might become  $PC_4$  in the second half. In order to test for that, we are then going to evaluate

the overlap of a PC in the first half with all the PCs of the second half, and retain the PC which has a maximal overlap.

$$|(PC_n^{(1)})'.PC_x^{(2)}|$$

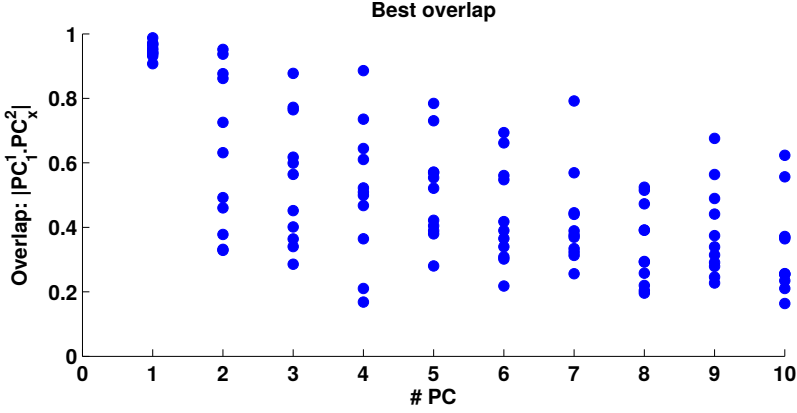


Figure 2.5: A: Overlap of the  $n^{th}$  principal component computed in the first half with PC computed in the second half of the recording which has a maximal overlap. Each dot represents one experiment. Parameters:  $t_c = 0.1s$  and  $N_{jitt} = 10$ .

We conclude from these results that the direction of the first principal component is very stable along the recordings.

### 2.2.4 Do all neurons participate in the competition ?

When thinking about how neural networks work, we expect that there is some underlying simplicity that allows its understanding. It could be that in a network, all neurons have a similar temporal pattern of firing (like in the synchronous state), or that each neuron says something at the right time, for example the letters of a word. In the competitive desynchronised state, we can think that neurons contribute to this latent variable by spiking at the right time.

The loadings are these coefficients that we attribute to each cell, and that we use to obtain the latent variable. We then proceeded to assess, using a non-parametric test (see methods 2.4.5.3), whether the firing rate of each cell was significantly correlated with the score of the population (computed omitting the firing of this particular cell).

As we can see in figure 2.6 A, the majority of the loadings of the first PC contribute to the score in a way that is statistically significant (in black) with respect to a situation in which the firing of the cell would have no temporal structure. This doesn't happen for PC's of higher order 2.6 B. We show this in the case of one experiment (c065a), but the results are similar for the other experiments.

We also plotted the correlation of the correlation matrix  $R^2$  (figure 2.7 B). Element  $R^2_{(i,j)}$  measures the correlations between the set of correlations of neuron  $i$  with all other neurons in the recording and set of correlations of neurons  $j$  with all other neurons. When the magnitude of  $R^2_{(i,j)}$  is large, this means that the subset of neurons with which neuron  $i$  is strongly correlated is similar to the subset of neurons with which neuron  $j$  is strongly correlated (i.e. neurons  $i$  and  $j$  have "the same friends"), although they might differ in sign. If the magnitude of  $R^2_{(i,j)}$  is close to zero, this means that knowing the cells with which neuron  $i$  is correlated does not allow us to predict the neurons with which neuron  $j$  is correlated. As we can see in the figure, even if correlations are weak for both pairs, these weak correlations are not statistical fluctuations, because the pattern of (weak) correlations of most neurons is fairly predictable from the pattern of correlation of other cells.

The conclusion of these two complementary analyses is that the majority of cells contribute significantly to the competition.



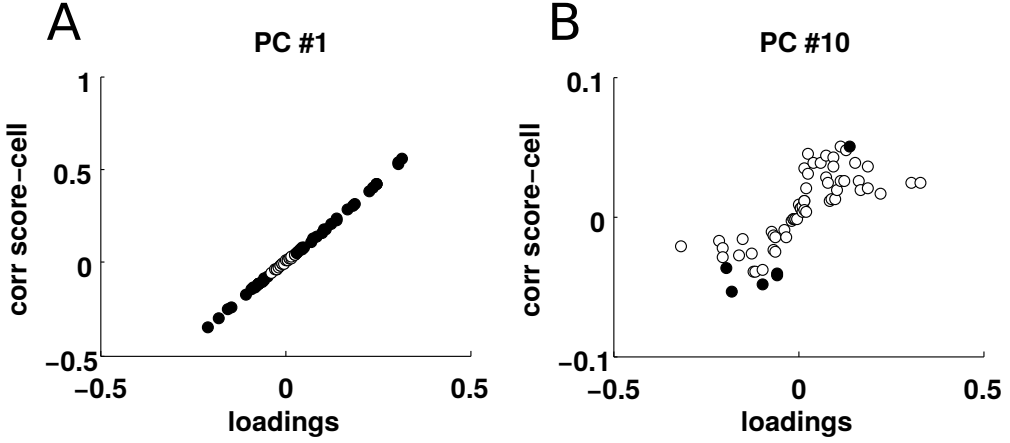


Figure 2.6: A: significance for different loadings of the first PC. In black (resp. white) the loadings for which the  $p$  value is lower (higher) than the inverse of the number of permutations. B: significance for the 10<sup>th</sup> PC. Experiment c065a. Parameters:  $t_c = 0.1s$  and  $N_{jitt} = 10$ , Number of permutations: 100.

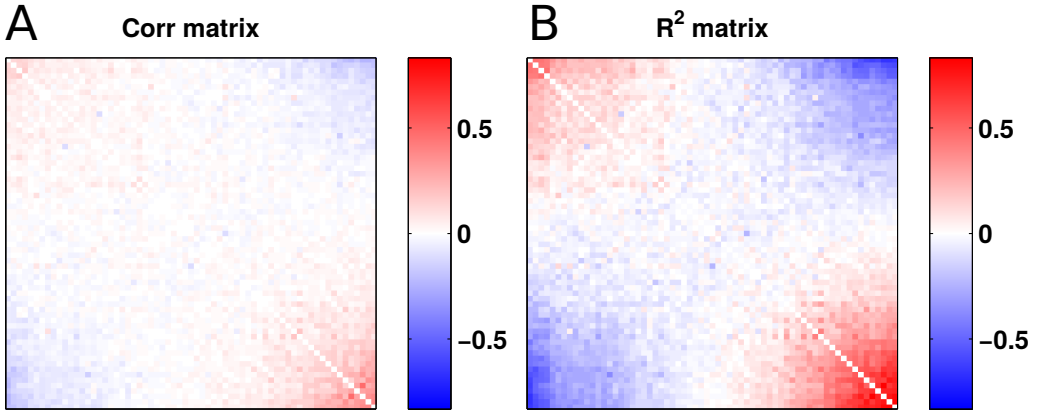


Figure 2.7: A: correlation matrix as in figure 2.2 E. B:  $R^2$  matrix of A. Experiment c065a. Parameters:  $t_c = 0.1s$  and  $N_{jitt} = 10$ .

## 2.2.5 Role of global fluctuations in competitive activity

### 2.2.5.1 Assessing the contribution of the mean activity to the population variance

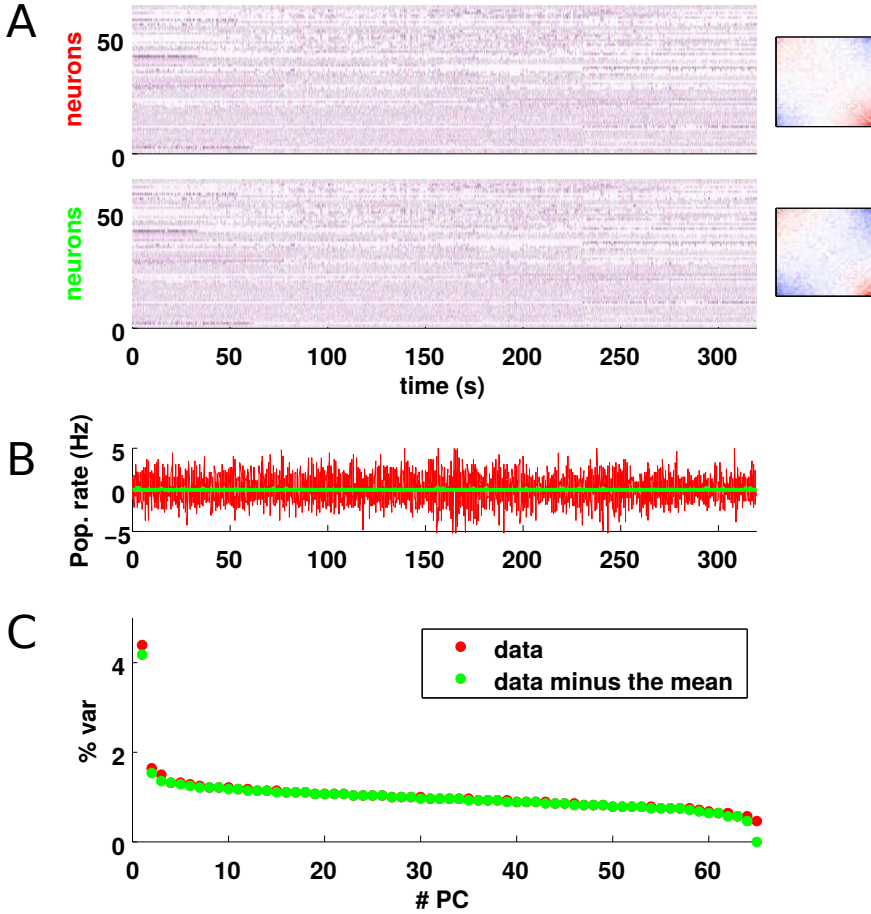


Figure 2.8: Necessity of the population rate fluctuations to explain competitive state. *A*: Count matrix of experiment c065a (red) and count matrix of surrogate data (green), to which we removed the instantaneous population rate. Right: corresponding correlation matrices. *B*: population rate of data (red) and of surrogate data. *C*: PCA of both data sets. Parameters:  $t_c = 0.1$  s,  $N_{jitt} = 10$ .

During the synchronous state, neurons go up and down together. The fluctuations at the level of the population rate explain a large fraction of the variance [109]. During the desynchronised state, the population rate remains roughly constant over time (figure 2.2 A). To assess how much these fluctuations contributed to the total variance of the population, we decided to remove these fluctuations and see if these fluctuations were **necessary** to explain the competitive dynamics. As  $D$  is our data matrix of dimension  $\langle T \rangle \times \langle N \rangle$ , the projection of  $D$  onto  $\vec{u}$ , a unitary vector of dimensions  $\langle N \rangle \times 1$ , is a time series of dimension  $\langle T \rangle \times 1$ :  $D \cdot \vec{u}$ . Then, projecting back this time series to the neural basis, results in a matrix of dimension  $\langle T \rangle \times \langle N \rangle$  and of rank one:  $D \cdot \vec{u} \cdot \vec{u}^t$ . We can then create a surrogate data set in which the fluctuations along the uniform direction are removed:

$$D^s = D - D \cdot \vec{u} \cdot \vec{u}^t$$

We depicted in red the original data set and the surrogate data set (green) in figure 2.8 A. In figure 2.8 B we show how indeed in the surrogate data set, the population rate -in green- is nil over time. When we apply PCA to the surrogate data set, we see that both its eigenspectrum and the correlation matrices of the original and of the surrogate data sets are very similar (figure 2.8 A and C). We then conclude that the fluctuations of the population rate are not necessary to explain the competitive dynamics. If to a first order in desynchronised state all neurons moved together, the activity in neural space would be in the direction of the uniform vector  $\vec{u}$ . However, in desynchronised state the  $PC_1$  has a shape that is symmetric around zero (figure 2.2 B and 2.3). This means that during desynchronised state the direction of the activity in neural space happens along a direction (given by the  $PC_1$ ), that is close to orthogonal to the uniform vector  $\vec{u}$ :

$$PC_1^t \cdot \vec{u} = \sum_i PC_1|_i \approx 0$$

### 2.2.5.2 Necessity and sufficiency of a given subspaces to explain the correlation structure

Lets briefly mention why PCA is a natural candidate to observe the most salient patterns of correlated activity in the data. Consider two variables  $n_1$  and  $n_2$ . We can plot all the observations of  $n_1$  and  $n_2$  in the neural

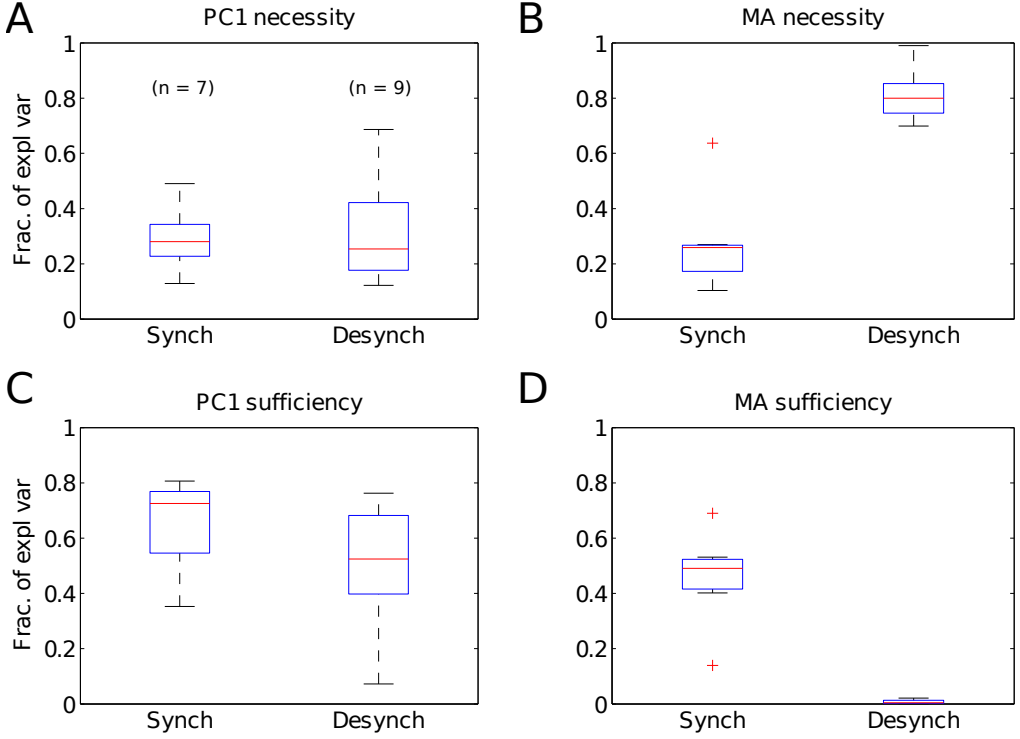
space of axis  $n_1$  and  $n_2$ :  $\{(n_1(t=0); n_2(t=0)), (n_1(t=1); n_2(t=1)), \dots\}$ . If we project orthogonally these observations onto a certain direction  $\vec{m} = \cos(\theta)e_1 + \sin(\theta)e_2$ , we obtain a time series  $\{p_m(t=0), p_m(t=1), p_m(t=2), \dots\}$ . We can compute the variance of this time series and obtain the projected variance. The first principal component gives precisely the direction that maximizes the projected variance. The second principal component (PC2) is the second direction that maximizes the variance in a direction orthogonal to PC1. For example if  $n_1$  and  $n_2$  form an ellipse on the neural space, the PC1 will point on the direction of the major axis.

We designed a method in order to assess how much fluctuations within a given subspace contribute to the full correlation structure of a data set (see methods 2.4.6). **Subspace** is a more general concept than direction, and means the portion of space generated by linear combinations of a set of directions  $\{\vec{v}_1, \dots, \vec{v}_p\}$ . The general idea of this methodology is the following: a subspace is considered to be **necessary** to explain the correlation structure when a surrogate database in which one has removed correlations exclusively along this subspace generates a correlation matrix which is dissimilar from the correlation matrix in the real data. In the example of figure 2.8, we did this in a rather coarse way by subtracting the projections along this dimension, instead of breaking the correlations along this particular subspace, and this is reflected in the fact that the surrogate data has a rank equal to  $N-1$  (see figure 2.8 C green). In a complementary manner, a subspace is **sufficient** to explain the correlation structure of the data, if a surrogate dataset in which one has removed correlations exclusively in the orthogonal complement of this subspace, generates a correlation matrix that is similar to the correlation matrix of the real data.

We then tested the necessity and the sufficiency of two different subspaces: PC1 and MA. Each component of the mean activity vector (MA) has the mean firing rate of the cell  $r$ , we normalize it so its norm is equal to one:  $MA = \frac{1}{\|MA\|} (r_1, r_2, \dots, r_N)^t$ .

In figure 2.9 A, we observe similar results for both states. When we destroy the correlations along the first PC, the fraction of explained variance is small. Thus because removing correlations exclusively along the first principal direction has a large and similar effect across both states, whereas removing correlations across all other directions has a relatively small effect also across both states, we conclude that the dimensionality of the data in both states is similarly low, at least as far as predicting the structure of the

correlation matrix is concerned (figure 2.9 C).



*Figure 2.9: Necessity and sufficiency of  $PC_1$  and mean activity. A and B: necessity of the  $PC_1$  and the mean activity vector MA. C and D: sufficiency of the  $PC_1$  and of the mean activity vector MA. Parameters:  $t_c = 0.1s$ ,  $N_{jitt} = 10$ ,  $N_{permutations} = 10$ . The y axis in the plots refers to the fraction of variance in the correlation matrix, not fraction of variance in the recording.*

In turn, when we examine the amount of variance explained by the mean activity vector, we note that when we break the correlations along the direction of the mean activity vector MA, we explain a very small proportion of variance in the synchronised state (figure 2.9 B), whereas we can still explain almost all the variance in the desynchronised state. We then conclude that the mean activity is necessary to explain the activity in the synchronised state, but not in the desynchronised state. When we break the correlations along the orthogonal complement of MA, we manage to explain a good fraction of the variance in the synchronised state, but we fail completely for

the desynchronised state (figure 2.9 D) . We conclude then that the MA vector is necessary and sufficient to explain the correlation structure in the synchronised state, but that it is not necessary nor sufficient to explain the correlations in the desynchronised state.

### 2.2.5.3 Comparison of directions between PC1 and population coupling vectors

In 2015, Okun et al. [109] argued that the population coupling vector was an interesting direction in neural space to observe the activity in the most "variable" -synchronised- states. The population coupling of a cell  $i$  with the population rate is:

$$PopC_i = \frac{1}{||f_i||} \int f_i(t) \sum_{j \neq i} (f_j(t) - \mu_j) . dt$$

$||f_i||$  is the firing rate of cell  $i$ , and  $f_i(t)$  is the continuous signal that results from convolving a spike train with a Gaussian kernel  $f$ .  $\mu_j$  is the mean activity of cell  $j$ . The population coupling can also be defined as a spike triggered population rate, and we can then define a lag, as in figure 2.10. The definition above refers to the zero lag case.

Using the original code of the population coding obligingly shared by M. Okun, we computed the population coupling (normalized by the constant  $\frac{1}{\sqrt{N}}$ ) for the usual experiment c065A, in both active and in inactive state (see figure 2.10 A and B). As we can see in B, in the active state, the population coupling is less important than in the inactive state.

We decided to compute a modified version of the population coupling, in which, instead of looking at the spike triggered population rate, we look at the spike triggered competitive latent variable:

$$PopC_i^{PC_1} = \frac{1}{||f_i||} \int f_i(t) \sum_{j \neq i} (f_j(t) - \mu_j) . l_j . dt$$

$l_j$  designates the  $j^{th}$  loading of the first principal component (of norm 1). As we can see in figure 2.10 C, in the inactive state, this quantity is very similar to the population coupling, whereas in the case of the active state, this direction explains more variance (figure 2.10 D).

Even if in figure B, we see that the population coupling is small, it still captures a little bit of variance. Given that the activity is approximately

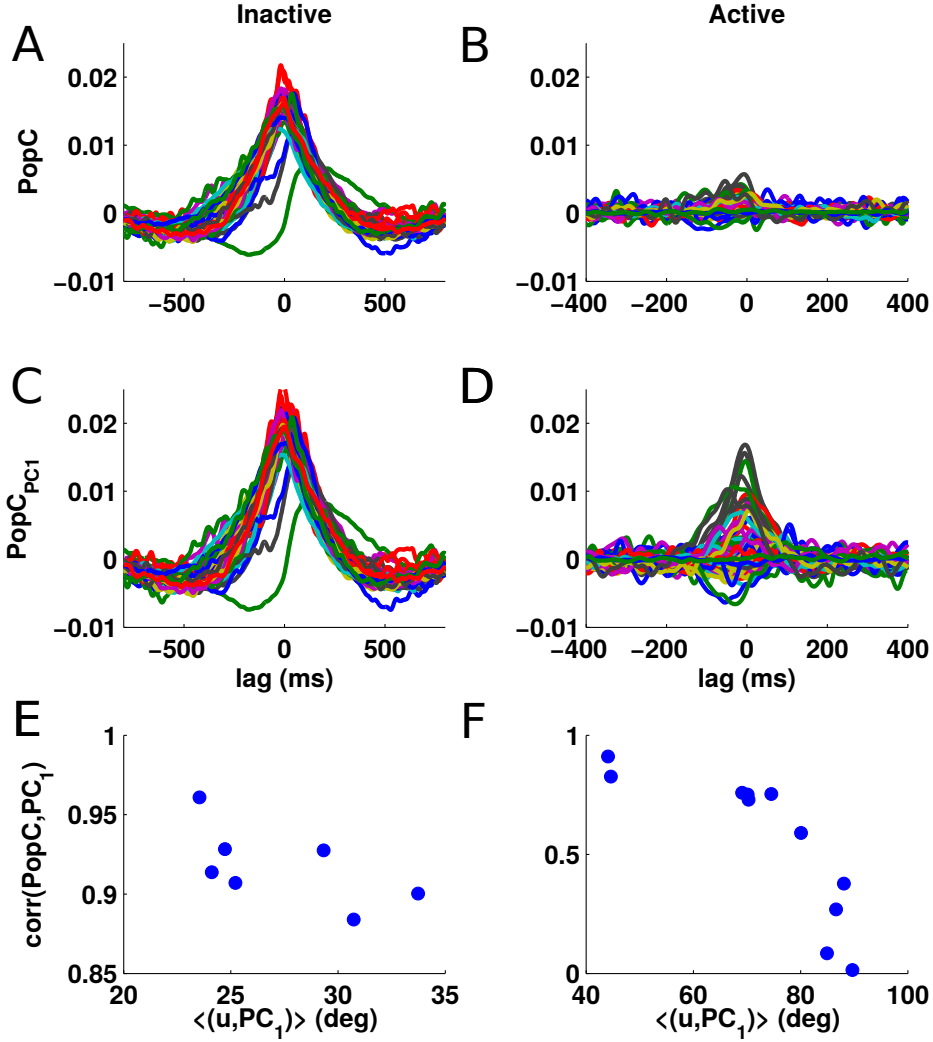


Figure 2.10: Relation between population coupling and PC1. A: Pop. coupling in inactive state. B: population coupling in active state. C (D): population coupling, weighted by PC1 loadings, in inact (resp. act). Each color designates a different neuron. E,(F): correlation between PC1 and population coupling as function of the angle between PC1 and uniform, in inactive and active state. Each dot represents an experiment. Parameters:  $t_c : 0.1s$ ,  $N_{jitt} : 10$ .

one-dimensional, this would be expected unless the uniform direction and the first principal component direction were perfectly orthogonal. In figure 2.10 F, we plot the similarity between the PC1 loading and the population coupling, as a function of the angle between the PC1 and the uniform. Consistent with this interpretation, in recordings where the PC1 is close to orthogonal to the uniform direction, PC1 loadings and population couplings are largely uncorrelated. During recordings in the synchronized state (figure 2.10 E), the first principal component is closely aligned with the uniform direction, and thus PC1 loadings and the population coupling are very strongly correlated. We conclude from this and our previous analyses, that whereas there is a substantial amount of temporal structure in spontaneous activity during desynchronized states, the nature of this structure is not particularly related to fluctuations in global population firing rate.

## 2.2.6 Single cell characteristics

### 2.2.6.1 Link between the $PC_1$ loadings and the firing rates

We saw before that nearly all cells contributed to the competitive activity. The projected activity of the population over the PC1, the score, is an fluctuating time series. The loadings, or components of the PC1, are scalar quantities affected to each neuron that quantify the degree to which cells contribute to this latent variable. We can therefore ask if there is some relationship between the loading and the firing rate of a cell.

In the desynchronised state, we have a complex relationship between the firing rate and the loadings of the PC1 component, that resembles an "U-shaped" relationship (see figure 2.11 A). Said otherwise, the firing increases as a the absolute value of the loading (see figure 2.11 B). To assess whether this relationship is statistically significant, we apply one non-parametric test to see the statistical significance between the absolute value of the loadings and the firing rate (see methods 2.4.5.1). When collapsing all the data from all the experiments, we obtain that this relationship is significant ( $p - value < 10^{-4}$ , 10000 surrogates). When we test the statistical significance on an experiment by experiment basis, we find that the relationship is significant in 6 out of 10 experiments ( $\alpha = 5\%$ ). The p-values are: (0.041, 0, 0, 0.1, 0.0047, 0.0077, 0.77, 0.46, 0.013, 0.45).

Here we would need patch-clamp experiments to observe the membrane



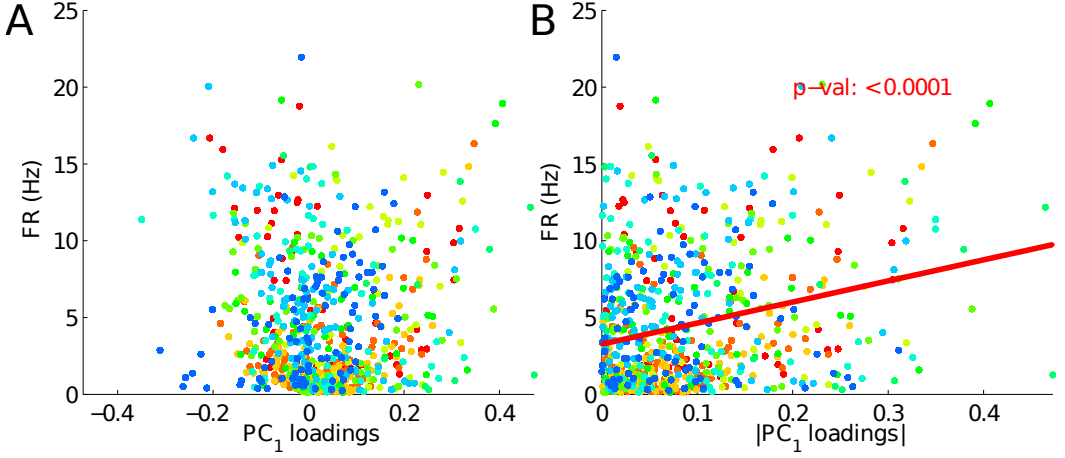


Figure 2.11: A: Firing rate versus  $PC_1$  loadings. B: Firing rate versus absolute value of the loadings. Each color represents an experiment. Parameters: ( $t_c = 0.1s$ ,  $N_{jitt} = 10$ ,  $N_{shuffles} = 10000$ ).

potential of the cells whose loading is small in absolute value, to disentangle between two possible scenarios:

a) it could be that the membrane potential of these neurons is correlated with the latent variable, but that the baseline potential is far away from the threshold, which would explain why the cell fires less. Even when the cell would spike, its correlation with the cells that spike more will be small.

b) Alternatively, it could be that the membrane potential is also far away from the threshold, but that it is not correlated with the latent variable, so that the firing rate is low and the spikes are weakly correlated with the latent variable.

The firing rate is the first characteristic one would look at when trying to relate a cell property to the loadings of the  $PC_1$ . The loadings give the contribution of this cell to the latent competitive variable. A second order characteristic that we would study and see if they correlate with the loadings are differences in the spiking pattern of neurons cells.

### 2.2.6.2 Link between the $PC_1$ loadings and the mean power of the spike trains

In figure 2.12 A, we can see that when we order the cells using the loadings of the first principal component (figure 2.12 B), we can see a difference in the

firing at the level of the raw data. On one extreme, cells seem more bursty, whereas on the other extreme cells seem to have a baseline firing rate on top of which cells increase or decrease their firing rate. Taking two example cells for the extremums of the competitive axis (in magenta and cyan), we compute the autocorrelogram of both in figure 2.12 C and D.

We then apply the Fourier transform (see methods 2.4.4) to all the spike trains in the recording to compute the power as a function of the frequency (figure 2.12 G). The power of each spike train is also the Fourier transform of the autocorrelogram. In figure 2.12 E and F we see the Fourier transform of our two example cells, we see that there seems to be a difference in power in the low frequencies. We then asked whether there was a consistent relationship between the mean power, in a frequency band of [0-4] Hz and the loadings of the first principal component. This relationship seems to be present (figure 2.12 H). To assess whether it is statistically significant, we did a non-parametric test fitting a linear model to the data (in red in figure 2.12 right), and comparing it with the slopes of surrogate data in which we randomly permuted the ordinates (see methods 2.4.5.1). Overall, the cells that are bursty have high power at low frequency, whereas the more regular spiking cells have a smaller mean power in the range 0-4 Hz. Across experiments, the relation is also significant in 6 out of 8 experiments. The p-values are: (0, 0.0032,  $10^{-4}$ , 0.4827, 0.0281, 0.0358, 0, 0.0815).

Previous work also points towards related features: Okun et al. [109] also found a correlation between population coupling and burstiness of the cells. Vinck et al. also observed a linear trend between firing rate and firing rate irregularity [154]. The relationship between the loading and the mean power is important, because it allows us to compare neurons across the one-dimensional competitive continuum across experiments. Since the direction of PC1 is arbitrary and therefore the sign of the loadings also, and also because the PC1 loadings are approximately symmetric around zero, it's not clear how to look for consistency across experiments. The fact that the autocorrelation of neurons at both ends seems to be systematically different allow us to disambiguate the sign of the PC1 loadings across experiments.

Although the loadings are approximately symmetric around zero, they are not perfectly symmetric (see figure 2.3). Given that to first order, the correlations are given by the product of their loadings, we can label as  $E_1$  the population that has bigger loadings, and adopt the sign for the PC1, in such a way that the bigger loadings (in absolute value) are positive.

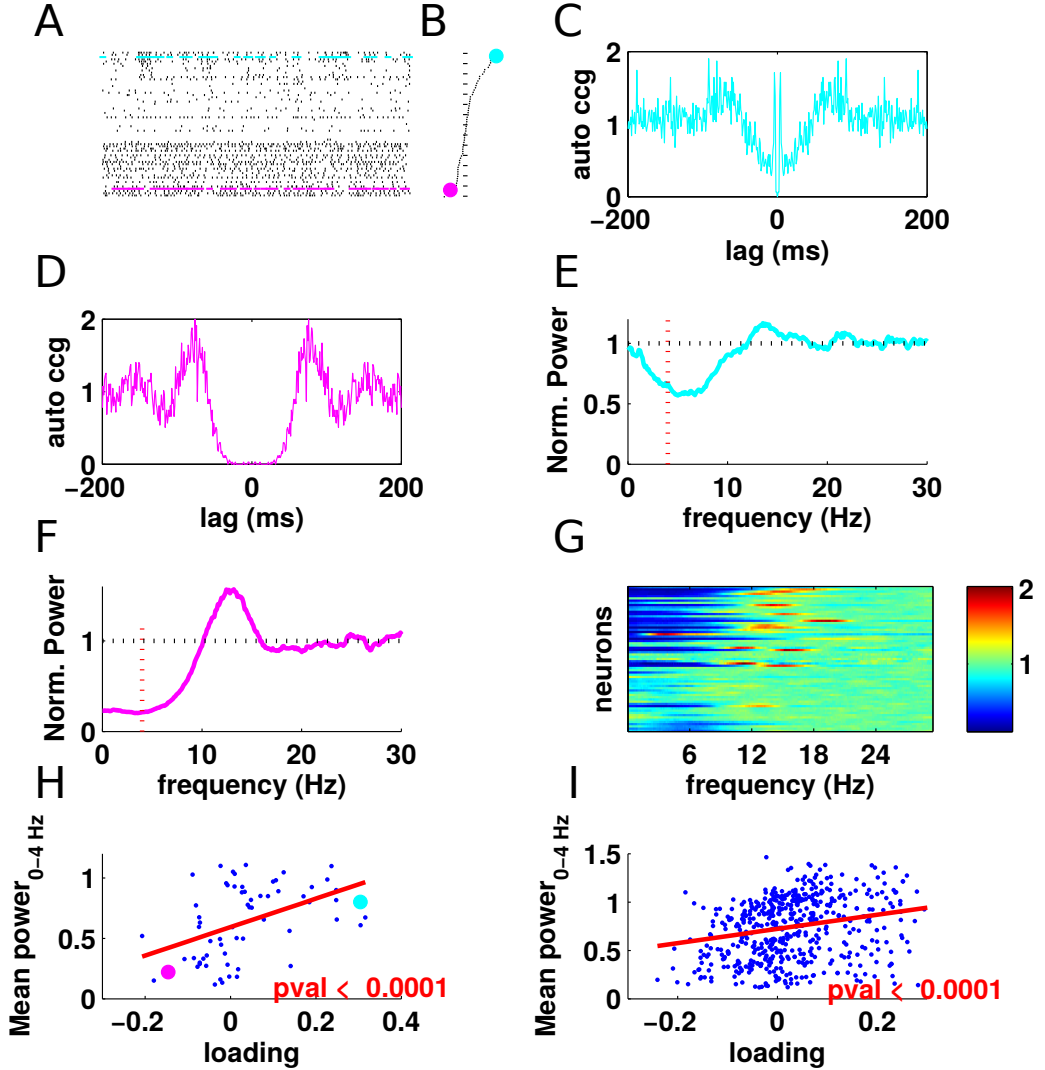


Figure 2.12: The mean power of single cell firing correlates with loading. A: raster plot of 065a (5 seconds). B: sorted PC<sub>1</sub> loadings. C,D: autocorrelogram of two neurons from the extremes of the competitive avis (in cyan and magenta). E, F: normalized power as function of the frequency for both cells. G: normalized power as a function of the frequency for all the neurons of the recording, the order is the same as in A and B. H: mean power (0-4 Hz) as function of the loading for the present experiment. Red line: linear fit. I: mean power (0-4 Hz) as function of the loading aggregating all the experiments. Parameters: ( $t_c = 0.1s$ ,  $N_{jitt} = 10$ ,  $N_{shuffles} = 10000$ ).

This would give us an alternative way of comparing the  $PC_1$  loadings across recordings.

### 2.2.7 Temporal characterisation of the competition

The competition can be seen in the data to the naked eye, and surprisingly this competition seems to happen at many time scales. In figure 2.13 we plotted in the middle a population raster with ten seconds of an experiment. In the top and in the bottom, we plotted a binned and z-scored matrix, ( $t_c = 0.1$  s and  $t_c = 1$ ). We plotted the **scores** on top and at the bottom of the count matrices. The scores are the projection of the network activity onto the first principal component, for every time point. As we see, both scores describe a alternation of the network at different time scales.

Remarkably, the PC is remarkably still at both times scales (figure 2.13 right). At left, we see how, keeping the order of the loadings of the  $PC_1$  computed at the larger time scale, in the  $PC_1$  computed at low time scale, doesn't alters too much the direction at which the PC points in the lower time scale. We also plot, in figure 2.14 the correlations at different time scales, and we can see the competition and also the fact that the PC direction is very still. In order to quantify this observation, we computed the PC basis at different time scales, using different time counts of  $t_c = 0.005, 0.025, 0.05, 0.1, 0.25, 0.5, 1$ s and a jitter window of five time counts. Therefore, in figure 2.15 we computed the overlap between a given PC at different time scales  $|(PC_i^{t_c^a})' \cdot PC_i^{t_c^b}|$ . The PC are vectors whose sign is undefined, that's why we consider only the absolute value of the scalar product.

In order to show that the correlations at big time scales are not due to correlations at smaller time scales, we did a control in which we maintain the correlation at a big time scale at which we compute the correlation, but we break the correlations at smaller time scales by redistributing randomly the spikes inside the (fixed) jitter windows of 1s. In the figure 2.16 we see that after applying the shuffling procedure to the present experiment 065a, the  $PC_1$  overlap is gone for time counts under 0.25 seconds. The time scale of the correlation is not given by the bin, but rather by the jitter window length  $t_c \cdot N_{jitt}$ .

To see if this result was consistent, in the figure 2.17, we averaged the overlaps of both the data and the control (figure 2.17 A, and B) across experiments. We present only the overlap for the first PC.

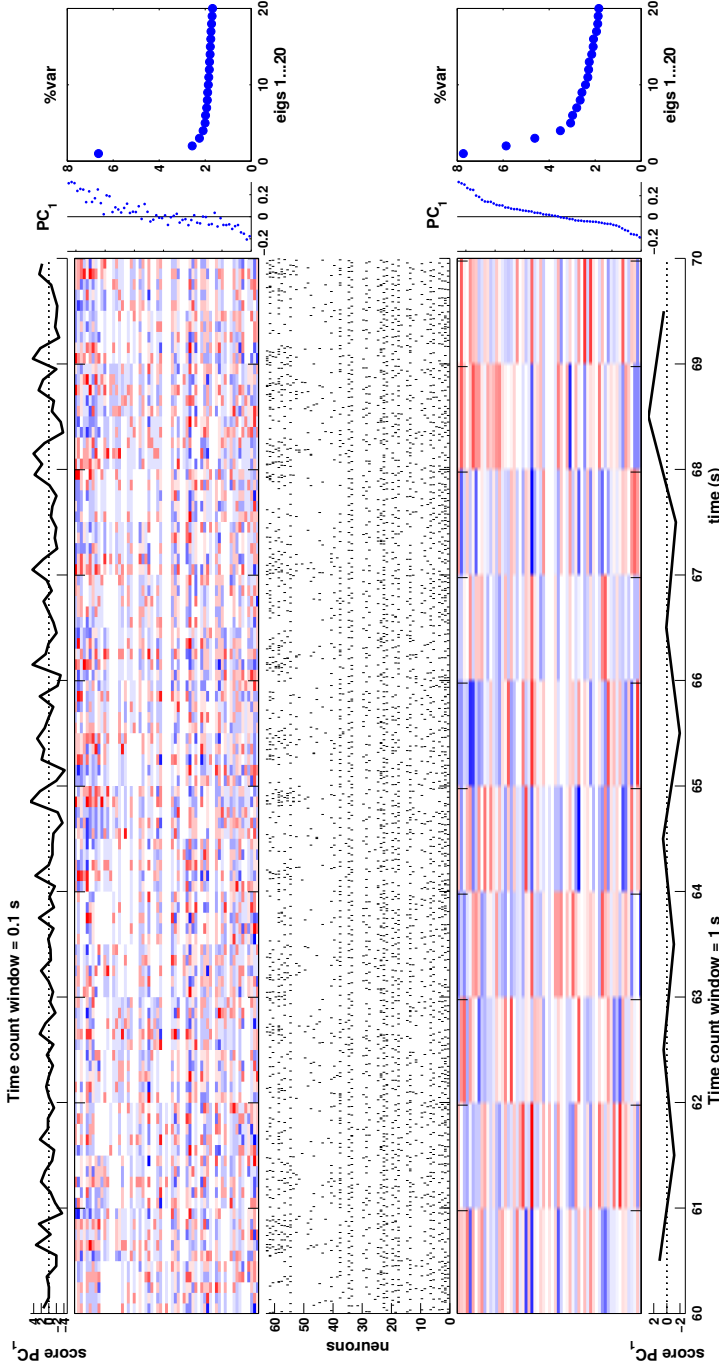


Figure 2.13: Competition happens at many time scales. Center: raw spike times, 10 seconds of experiment "c065a". Top and down: count matrix with fixed time windows and local means, with the same order of neurons ( $N_{\text{jitt}} = 5$ ) see methods 2.4.2). We look at the covariances at two isolates time scales: the time count window is 0.1s and 1s. Upper and downer: projected activity into the  $PC_1$  at the respective time scale. Left (up, down) percentage of explained variance.

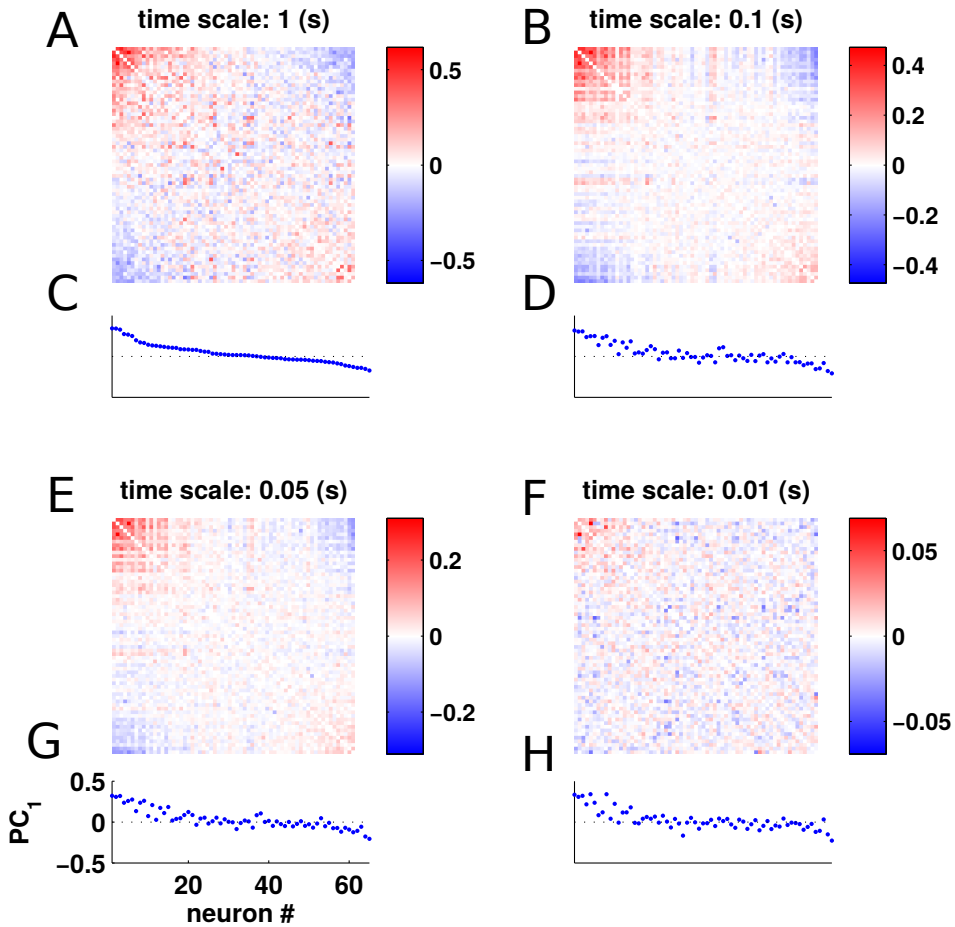


Figure 2.14: Correlation matrices at different time scales, using a different time bin for each matrix (A, B, E, F). We also plot the corresponding  $PC_1$  (C, D, G, H). The order of both the matrices and of the  $PC_1$  at different time scales is the order of the  $PC_1$  at the higher time scale  $t_c = 1s$ , in A. Experiment "c065a". Parameters:  $t_c = 1, 0.1, 0.05, 0.01s$ ,  $N_{jitt} = 10$ .

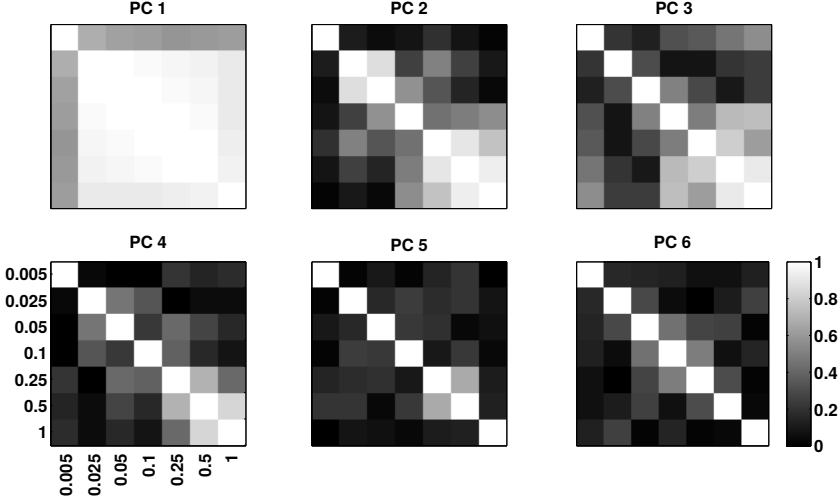


Figure 2.15: Alignment of the different PCs at different time scales for the experiment "c065a".

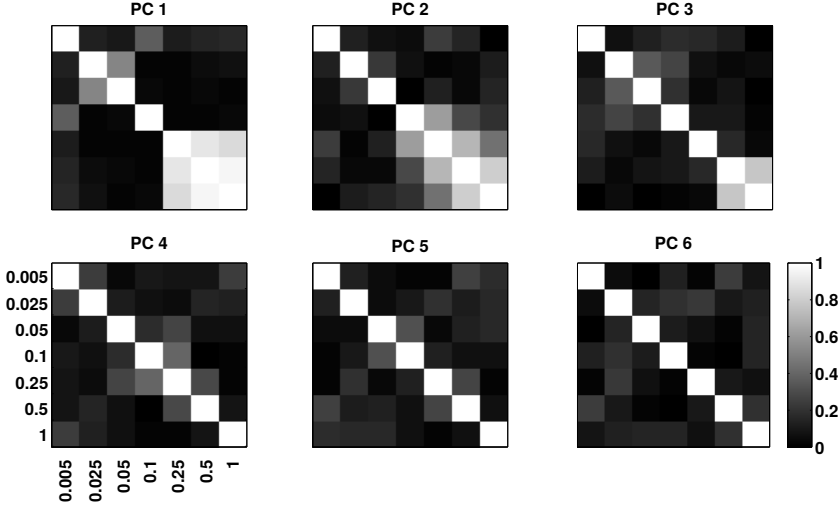


Figure 2.16: Alignment of the different PC at different time scales, for the experiment "c065a", after having destroyed the correlations at smaller times scales. Parameters  $t_c = 0.005, 0.025, 0.05, 0.1, 0.25, 0.5, 1s$ .  $N_{jitt} = 10$ .

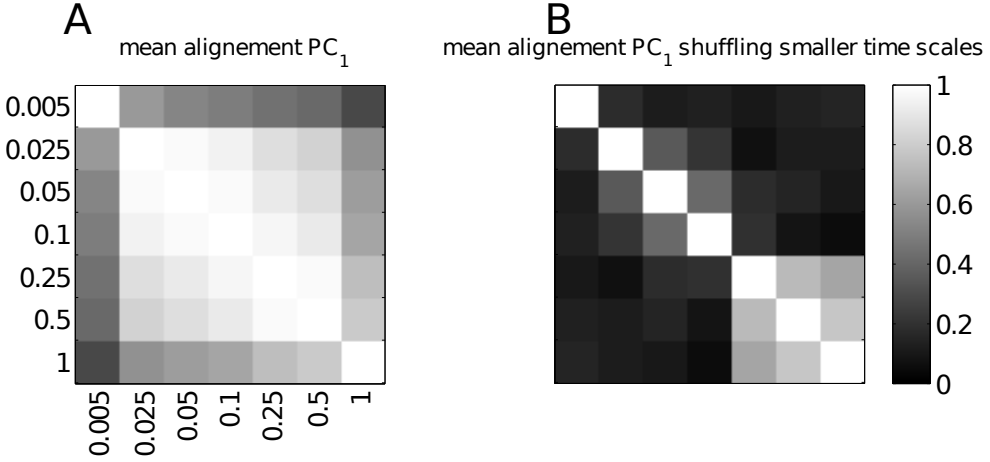


Figure 2.17: Summary statistic. Mean alignment across experiments of the different PC at different time scales. A: original data. B: after having destroyed the correlations at smaller times scales.

To conclude, we can then see that the correlation indeed happens at many time scales, and that this effect is consistent across experiments.

## 2.2.8 Spatial characterisation of the competition

### 2.2.8.1 Competition is local at the level of single shanks

The sorting of the correlation matrix reveals that cells can be organized along a continuum with functionally distinct properties at either end. In this paragraph, we will deal with the spatial aspect of the competition, in order to figure out whether there are two segregated clusters of neurons that compete, or if neurons from different groups are spatially intermingled.

We assessed how the PC1 loading of neurons depended on the shank (out of eight) of the silicon probe from which they were recorded. Two scenarios are possible (see figure 2.18): either we see in the data two clusters of cells with respectively positive (and negative) loadings, or we see competition in a same shank. What we see when we do the experiment, is that at the scale of the inter-shank distance, the competition is local (see figure 2.18 right).

To quantify the observation of figure 2.19, we elaborated a statistical test to distinguish between the scenarios C and D of figure 2.18, i.e. whether we



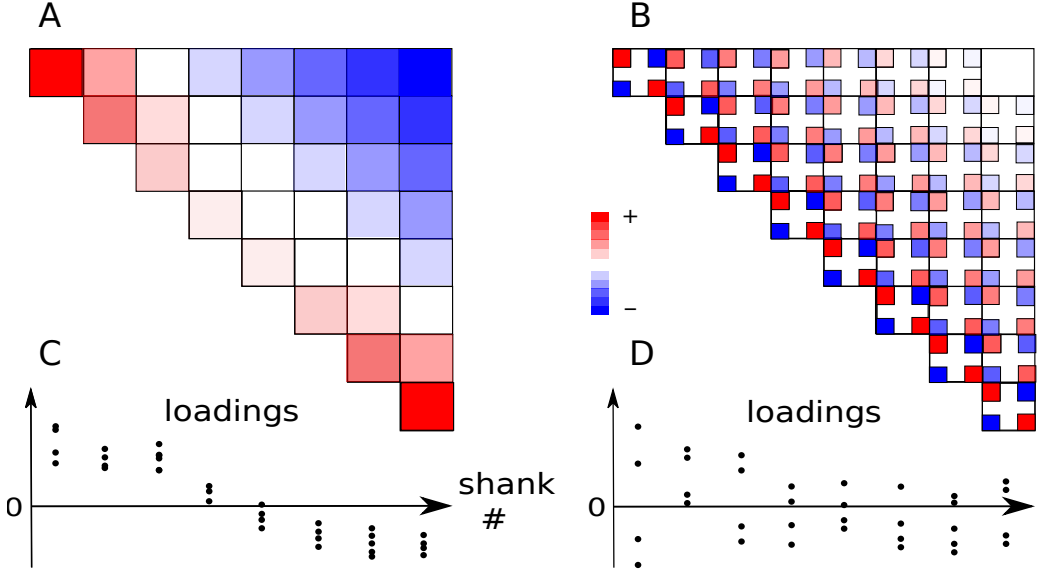
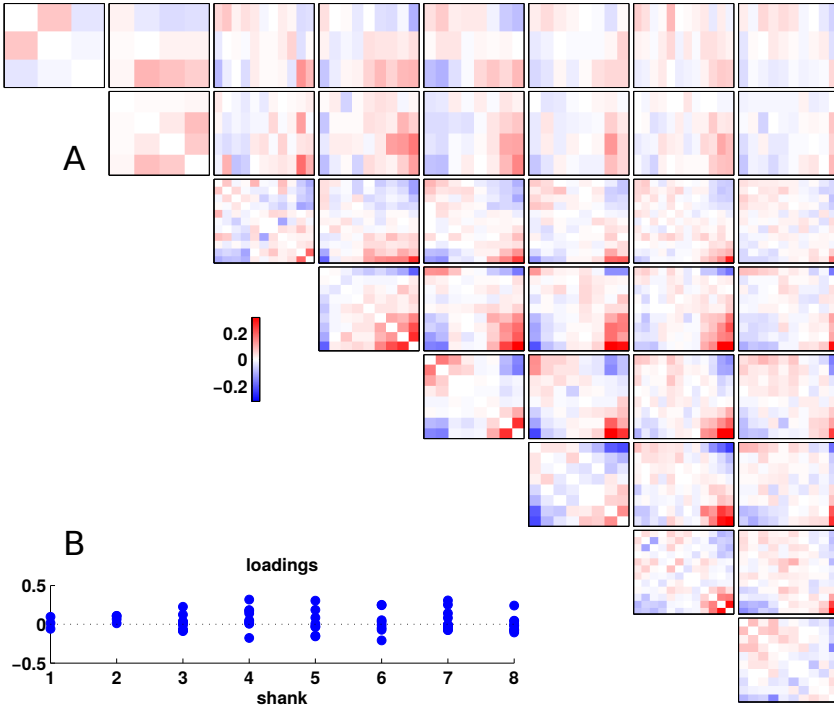


Figure 2.18: *A and B: schematics of correlation matrices with alternative scenarios of competition. The loadings are attributed to their respective shanks (C in the case of A and D of B). Depending on the sign of the loadings in each shank, we will see spatially segregated competition (A) or local competition (B).*

would see two spatially segregated groups of positive and negative loadings, in every shank.

We reasoned that if neurons with positive (negative) PC1 loadings are recorded in different shanks (figure 2.20 A), then if we group all shanks into two groups (those with median positive (MP) loadings, and those with median negative loadings (MN) (figure 2.20 B), the loadings in these two groups should have the same sign and opposite signs across groups. Thus the products of within-group loadings would be positive but those of across group loadings would be negative, and thus have non-overlapping distributions. If on the other hand, each shank contains both positive and negative loadings, the group of MN shanks will also contain sign-mixed loadings and the distribution of products of within group and across group loading will be highly overlapping (figure 2.20 C). We quantified the degree of overlap between these two distributions using the Area Under the Curve (AUC) of the corresponding Receiver operating characteristic ((ROC) curve for these



*Figure 2.19: Competition is local at the level of single shanks. We affected the cells to the shank they were recorded from and then sorted the correlations locally following the order of the  $PC_1$  loading of each cell (A). In B, we simply affected the loadings of the  $PC_1$  to their corresponding shank.*

two distributions. As we can see in figure 2.20 D-E, the AUC is very close to 0.5 for all experiments where we could conduct this analysis (sufficient neurons recorded in each shank).

In synthesis, we showed that the competition is not spatially segregated, but rather local: at the scale of single shanks, the loadings have as much positive as negative loadings, inducing both positive and negative correlations.

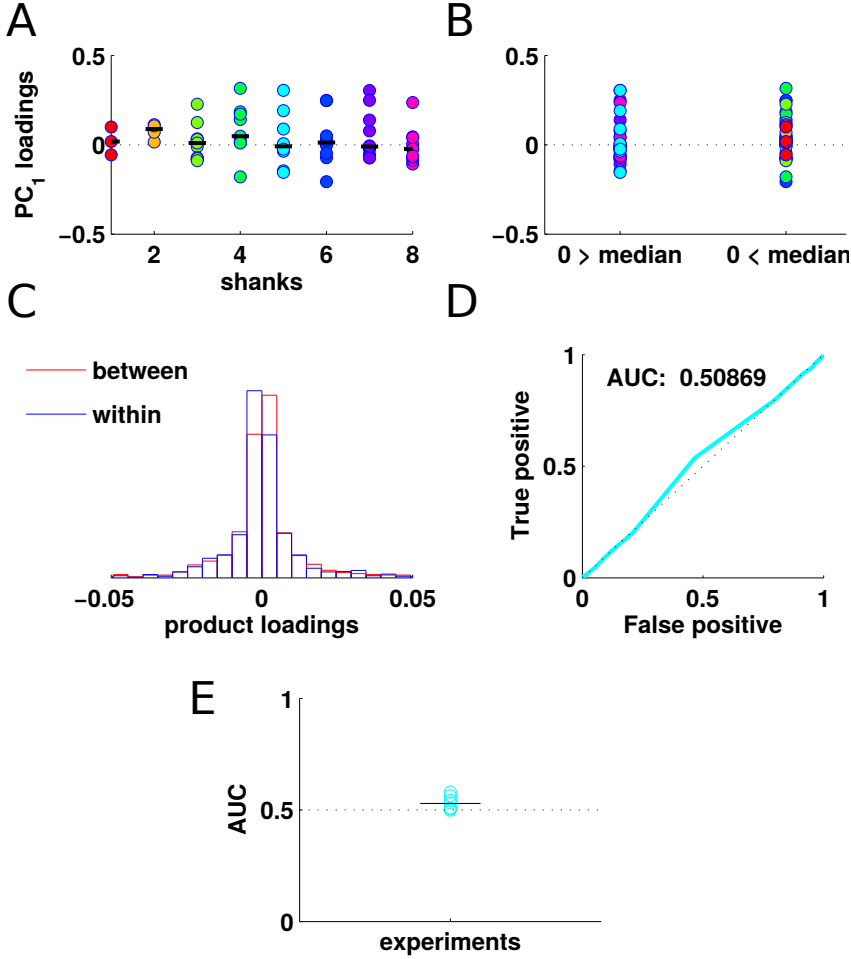


Figure 2.20: Quantifying the spatial heterogeneity in the sign of the neurons loadings. A: loadings assigned to the shanks the neurons were recorded from. Black bar: median. B: shanks loadings are attributed to two different groups: whether they median is negative (MN) or positive (MP). C: histograms of the product of the loadings between groups (red), on within groups. D: receiver operating characteristic curve, giving the true positive rate as function of the false positive rate. The bigger the overlap between the two distributions, the closer the area under the curve AUC is to 0.5. E: AUC for each experiment. Parameters: experiment c065a.  $t_c = 1, 0.1, 0.05, 0.01s$ ,  $N_{jitt} = 10$ .

### 2.2.8.2 Evolution of the spread of the correlation with the distance

Renart et al [123] showed in their supplementary information that during the desynchronised state the mean pairwise correlation of neurons stays constant with the inter-shank distance between these cells. We redid and confirmed those analysis (not shown).

Another point that we can test is whether competition is locally higher at a certain point in space and then decreases with the distance between shanks. This would be reflected in the spread of the loadings distribution. For each experiment, we assign the loadings to the shanks they were recorded from, and we compute, for each shank, the standard deviation of the loadings. Then, for each recording we find the location that has maximal standard deviation, and align this location across experiments. In figure 2.21 A, we can see with the naked eye that for almost all experiments (colors), the standard deviation of the loadings decrease monotonically with respect to their maximum. In order to test this ordered decrease with respect to the maximum, we devised a statistical test that would help us distinguish between the fits of a parabola done to the data and the fits of a parabola to surrogate data in which the spatial structure (excepting the maximum) would have been randomized.

With this non-parametric test (see methods 2.4.5.2), we can then see if the parabolic fit of the normalized and aggregated data is inside the envelope of the fits on the surrogate data (figure 2.21, B), and even compute the p-value, (figure 2.21, C). We see that the decrease is statistically significant in almost all the locations.

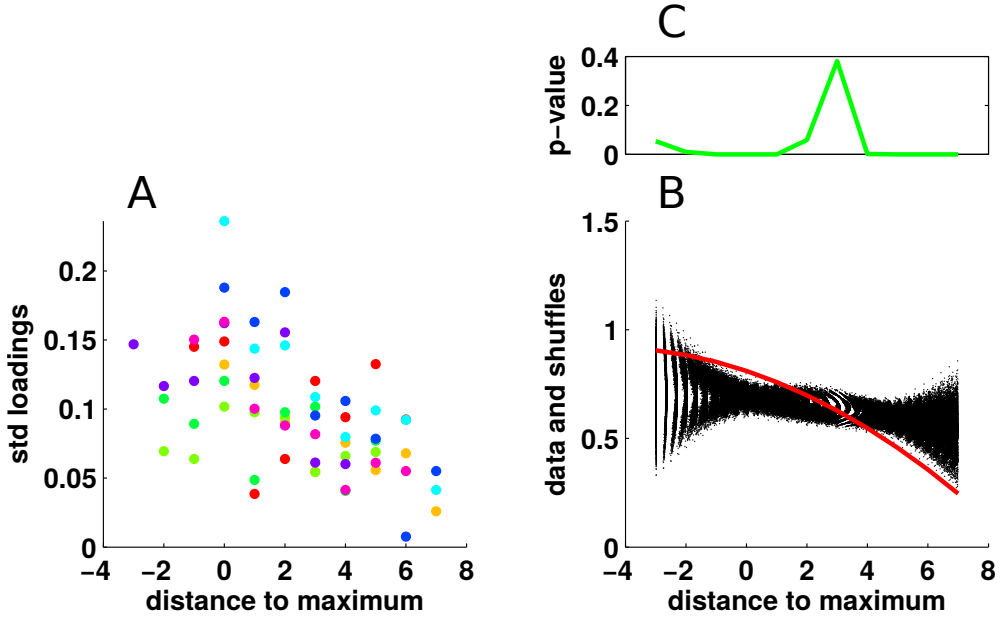


Figure 2.21: Evolution of the the standard deviation of the loadings with the distance. *A*: raw data: each color represents an experiment, each dot represent the standard deviation of the loadings in a shank. We aligned the data with respect to the shank which had maximal standard deviation, and flipped if necessary the orientation in order to have better statistical power. *B*: reassigning randomly the standard deviations (omitting the maximum), we create 10000 surrogates. For each surrogate we fit a parabola, and we compare the parabola fit on the data with the distribution of parabolas fitted on the surrogate data. *C*: p-value of the data as a function of the distance to the maximum.

## 2.3 Discussion

### 2.3.1 Summary

In this chapter, we report the finding of competitive activity during the desynchronised state of spontaneous activity. We find that the population activity is near one dimensional and that we can split the population of neurons in halves in which we observe coherent activity. The two populations alternate their firing in such a way that the overall firing rate remains constant.

We showed that the amplified direction is very still across the recording and that almost all the neurons contribute significantly to the competition.

Neurons in each of the two groups have different firing patterns, one being more bursty and the other being more regular. The firing patterns of each group show a different signature in the frequency domain: one has more power in the frequency band 0-4 Hz. Also, we found that the PC1 loadings correlate with the mean power in the band 0-4 Hz.

We characterised temporally the competition, and found that surprisingly, it happens at many time scales. When characterizing the spatial aspect of the competition, we found that this competition happens locally at the spatial scale of a recording shank. Also, we saw that the competition is stronger locally and then decrease the spread of the loadings with the distance.

### 2.3.2 Temporal invariance of the competition

The temporal invariance (see [2.2.7](#)) of the competition is truly a puzzling effect that its hard to know how it might be generated. These extraordinary complex scale-free effects appear in nature everywhere, and are the product of very simple rules that aggregate or affect elements, and also recursively affect the aggregated ensemble of these elements. One aspect that we showed in the previous chapter is that competition seems to be stronger in certain points and then decrease with the distance (see [2.2.8.2](#)), so in more distant places, the loadings are weaker, (and therefore the local positive feedback is smaller, see next chapter) , and then the competition locally could be faster and have lower power. We can therefore hypothesise that the temporal invariant aspect of the competition might result

from a linear superposition of spatially-propagating competitive effects of different time scales. At the network level the latent variable (the score) seems like a Weierstrass function, which is precisely a fundamental frequency on top of which are added harmonics of lower frequency and power ( $f(t) = \sum_{n=0}^{+\infty} a^n \cos(b^n \pi t)$ ,  $0 < a < 1$ ).

### 2.3.3 Physiology, connectivity and competition

We do not know for a fact if the two competing groups correspond to two different cell types. We hope that further studies get to identify if these two groups could be identified to the so called cortico-striatal and cortico-spinal pyramidal neurons [81]. Kiritani and colleagues show an asymmetric connectivity between the two groups but in a way that differs from the model: cortico-striatal to cortico-striatal show strong connections and cortico-spinal to cortico-spinal show strong projections but less than cortico-striatal to cortico-striatal. Also, they find asymmetry in the connectivity between these two groups of cells: cortico-striatal project strongly to cortico-spinal, but the connection between cortico-spinal and cortico-striatal is very weak. This is not in the line with the model, because in the model the population that connects strongly to itself also receives stronger projections than the projections it sends to the other competing group.

Another very important factor that might play a role in the competition, as shown by Kampa and colleagues [76], is that in layer 5 there are sub-networks of preferentially connected neurons, in which each sub-network tends to receive more shared input from cells in layer 2/3. This might have a functional relevance at the level of the whole column, because in turn, pairs of cells in layer 2/3 tend to receive more shared inputs from layer 4 [164].

Also, Rock and Apicella [124], find in A1 two separate pathways from inter-hemispheric projections into layer five: in the first pathway, excitatory neurons project preferentially to fast spiking inter-neurons, which in turn project mainly to cortico-cortical neurons. In the second pathway inter-hemispheric projections project directly to cortico-collicular neurons of layer 5. These facts point at the fact that the simpler hypothesis of anti-correlated input into L5 is also possible. Nevertheless, it is also compatible of the asymmetric inhibition predicted by the TCA model.

When we examine the temporal properties of the firing we see that in-

deed at the extremes of the competitive axis, the firing properties are very different: on one side, neurons are more bursty, and on the other side of the competitive axis, neurons are more regular. We showed that the mean power in a frequency band between 0 and 4 Hz, correlates in a significant way with the loadings of the PC. This is a good physiological landmark that might serve for further investigation. In barrel cortex, Agmon and Connors [1] identified -back in 1992- four kinds of different cells: fast spiking, intrinsic bursting, regular spiking "1" and "2". In layer 5, the most represented are the regular spiking "1" and "2" and the intrinsic bursting ones. Consistent with the literature, Rock and Apicella [124] identified recently, in layer five of auditory cortex, two kinds of cells with different morphologies and electro-physiological properties: cortico-collicular, which have extensive dendritic arborisation, bursty firing and high resting membrane potential. The other kind of cell they found, the cortico-cortical neurons, had more regular spiking, higher input resistance than the cortico-collicular, and a larger action potential half-width than cortico-collicular. However, with regard to the identity and the synaptic organisation of the inhibitory neurons mediating lateral inhibition, not much is known [166].

All these facts point to the fact that the connectivity in layer five is highly complex, and that we still need more studies about the connectivity. If it turned out that the cells with positive loadings correspond to a defined cell type, we still would like to know how to explain the gradation in the loadings from a physiological point of view, quantifying for instance the connectivity strengths, as in [109].

### **2.3.4 Functional significance of competitive dynamics during spontaneous activity**

#### **2.3.4.1 Change of directions of spontaneous activity during state changes**

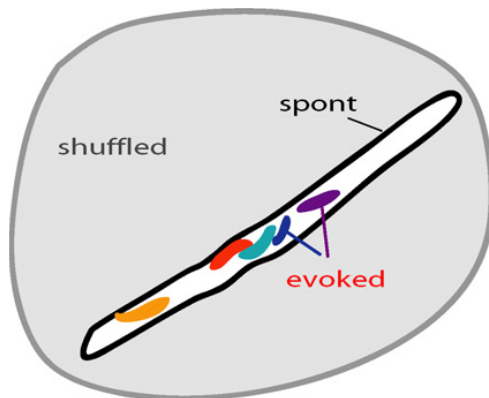
Okun et al [109] showed that the population coupling is the relevant direction to look at when observing data from the most synchronised states. However, in the most desynchronised states, the population coupling doesn't explain a large fraction of variance. There are two situations consistent with this finding: one that there is nothing to explain because during the desynchronised state the activity is temporally unstructured. The other one is that there is



something to explain, but the population coupling is not the right method to reveal it. Our results are coherent with this second scenario, and show that there is a similarly low dimensional pattern that drives variability in the desynchronised state, and that modulates the activity of all the neurons in a similar manner.

### 2.3.4.2 Directions of spontaneous and evoked activity

Luczak et al [89] advance the theory that spontaneous and evoked activity are confined in the same subspace of the neural space, as shown in the sketch of figure 2.22.



*Figure 2.22: Directions of spontaneous and evoked activity in neural space. Adapted from Luczak [89].*

In a more recent study, Luczak et al [90], presented repeatedly pure tones for one second, and observed that the probability of eliciting an up state increased significantly after the stimulus presentation. Across trials, the effect of stochastic up and down events averages out, so the population rate over trials seems constant and lets appear a transient increase in the population rate (see figure 2.23). This shows that on average, during evoked activity, the direction of the population activity is the same as the direction during synchronous state of spontaneous activity.

Okun et al [109] and Luczak et al [90], provide two different features to characterise how individual neurons are coupled with the population rate both in spontaneous activity and in presence of a stimulus: how reliably

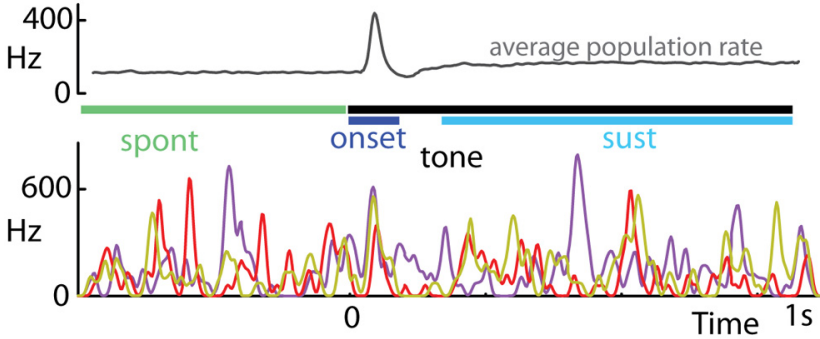


Figure 2.23: On top: mean population response to a one second pure tone averaged across trials. Bottom: example of fluctuations of the population rate in three single trials. Adapted from [90].

they are coupled to it (quantified by the population coupling, which is the ordinate of the cross-correlogram of one neuron with the population rate at zero lag), and with what delay they are coupled with the population rate ( $\mu_{cc}$  is the center of mass of the cross-correlogram of one neuron with the population rate, and is homogeneous with time).

#### 2.3.4.3 Luczak's gating hypothesis of communication between brain areas

The increase of the population rate when a stimulus is presented (figure 2.23), is interpreted by Luczak as a possible mechanism with which the onset of sensory inputs is gated, and which signals to upstream areas of the cortex that a signal arrived.

As we said in the introductory chapter, if this sensory gating hypothesis is true, we need to find an explanation for how upstream areas distinguish which up phases carry a stimulus. A simple hypothesis in the framework of the packet based communication theory of the cortex [92], is that the information -and its absence- is coded in the fine structure of the spikes sequence, which is coherent with the findings of Luczak's lab [14], [89], in particular that the sequence of spiking in the "packets" is modified with learning and that is dependent on plasticity mechanisms.

We will now examine a preliminary experiment during desynchronised activity, that shows that the uniform direction might not be the only one relevant for the evoked activity.

### **2.3.4.4 Evoked and spontaneous activity during the desynchronised state**

In a single preliminary experiment realized during desynchronised activity, click evoked responses were recorded (figure 2.24). As in the Luczak et al. paper [90], the mean activity (in black) has a transient increase (very visible at the level of the PSTH across experiments). However, soon after the rise of the mean (or uniform) activity, the direction of the activity rotates and aligns with the competitive direction. To see this, we have to split the population in two halves corresponding to the two competitive populations, and average the responses of each of the groups (figure 2.24 A). This population effect happens on a single trial basis (left column), when we average the two populations, and it is consistent across trials (right column). In the next section we propose an hypothesis of how to interpret this observation in the context of the stimulus gating during the desynchronised state.

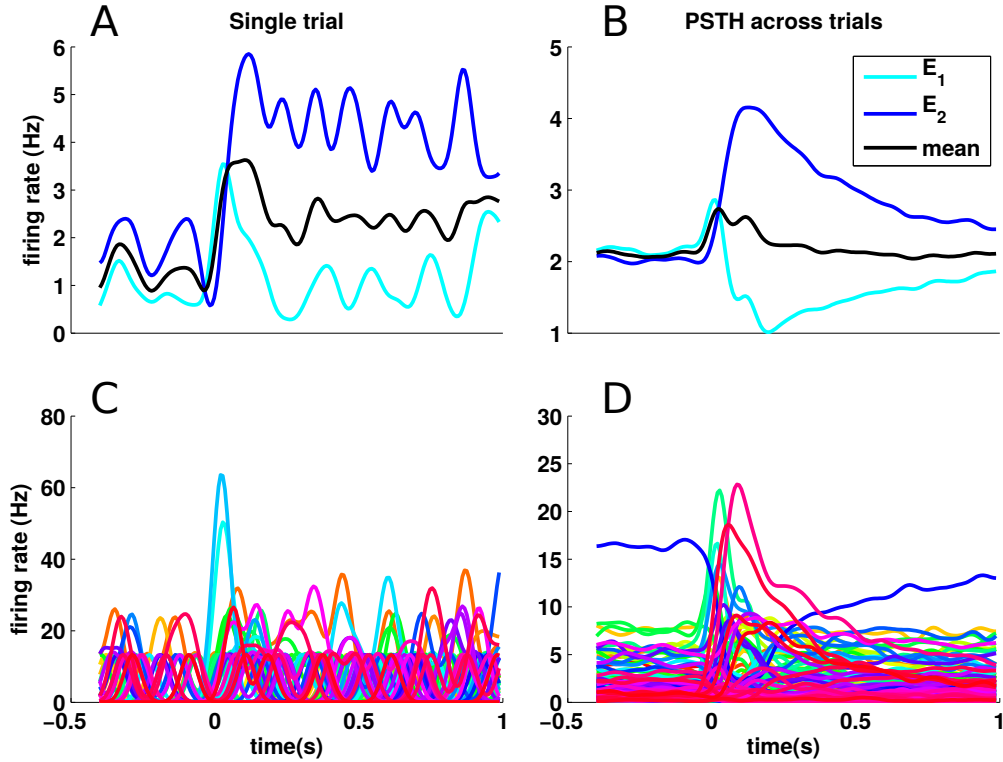
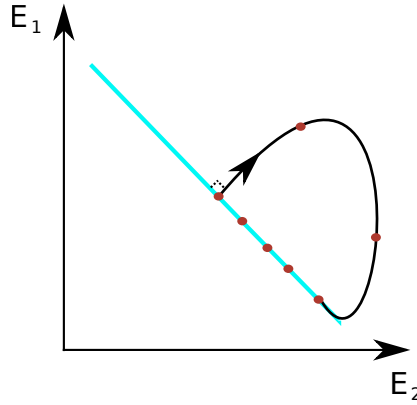


Figure 2.24: Biphasic population response to a click. *A,C* : single trial, *B,D*: average populations response to a click. *A,B*: we average separately the neurons with positive and negative  $PC_1$  loading, and compute the mean. *C,D*: single cell responses of all population neurons. Each color represent a different neuron. Each spike train is convolved with a Gaussian kernel of standard deviation 30 ms.

#### 2.3.4.5 The null space hypothesis of evoked activity during the desynchronised state

In the introductory chapter, we went over the different ways by which other brain areas could distinguish between what is signal and what is noise. One of the advanced hypothesis had to do with the direction of the activity. Kaufman et al. [78] advanced the hypothesis that as long as a linear combination of the activity was constant in an ensemble of pre-synaptic regions, the post-synaptic neurons wouldn't notice the variation in the sum of its inputs. As during the desynchronised state the population rate is maintained constant

dynamically (by being in a competitive state), when a stimulus comes, it elicits a change in the direction of the activity, aligning transiently with the uniform mode (which is orthogonal to the competitive mode). The uniform activity is damped fast, but the activity of both populations stays higher than average, and is damped until it attains the levels of the spontaneous activity (see figure 2.25).



*Figure 2.25: This phase diagram represents the transient dynamics during evoked activity as seen during one preliminary experiment. The axes represent the population firing rate of the two competitive populations  $E_1$  and  $E_2$ . The dots represent equally spaced time points, and highlight that the portion of the transient over which the mean activity increases and decreases (black) happens much faster than the portion of the transient over which the activity is competitive. The initial direction of the activity following a stimulation is orthogonal to the ongoing activity, and realigns with the spontaneous activity in a late phase.*

We then reuse the hypothesis of Luczak et al [90], that the initial transient activity along the uniform direction is useful to propagate activity along cortical areas. However, as we see in figure 2.24 and in figure 2.25, the activity along the uniform direction is quenched fast. When we see how does the population activity of  $E_1$  and  $E_2$  evolve, we see from this preliminary experiment, that  $E_1$  and  $E_2$  they stay up for much longer time (figure A and B). Therefore, soon after the first global perturbation, the circuits switch into a competitive mode. This could be useful for isolating functionally cortical areas of the hierarchy, because if these two populations project to a same

upstream region, this region will not notice the change in the activity in the downstream region.

One of the possible draws of this hypothesis, pointed by Paul Bush, is that if the two competing populations are the two distinct classes of neurons that have been identified in layer five, then these neurons should also project to the same brain region. Some studies point at the fact that the different classes of neurons in layer V are intermingled, have asymmetric connectivity and project to different brain areas. These neurons are called cortico-spinal and cortico-striatal [81]. Although this point seems important, we still don't know much about the competitive populations.

### 2.3.5 Low dimensional dynamics in cortex

One of the questions that arises is, why would cortical dynamics be low dimensional ?

One trivial, but possible answer would be that there is a simple underlying process, like a directional flow, or a steady process. Take for example the water molecules in a river, they move with respect to each other in very complicated ways, but overall, they move coherently with the flow. In the case of spontaneous activity during active state, it could be that what we observe is some kind of resting mode, in which the network propagates the activity for homeostatic reasons, but at the same time this activity doesn't harm the sudden information processing needs. This is compatible with the previous view of the null space hypothesis.

However, the null space hypothesis do not preclude that the spontaneous activity over the null space has a functional role. Regarding sensory processing, it could be that the spontaneous activity moves along the null space in order to generate some attentional or expectation effect that could improve discrimination and reaction time [42], [44], [130]. In what concerns motor activity, the null space hypothesis was coined by Kaufman et al [78], in order to explain preparatory activity in premotor cortex.

In fact, in the majority of computational frameworks in which recurrent networks perform computations (attractors [68], predictive coding [17], reservoir computing [21], [141], linear dynamical systems [46], [52]) the connectivity matrices have low rank and generate low dimensional dynamics also (see next chapter). Hennequin [65] shows however how the high dimensionality, even in a stable linear dynamical system, can be used as an

interesting substrate to generate very rich dynamics.

In our case, the hypothesis of Ganguli et al [46] seems a parsimonious in-between a purely passive process and a functional process, because it leaves the door open for both: the circuit wires itself in such a way that it put constraints on the dynamics. Certain dynamical quantities -like the population firing rate- are held constant, in spite of the large heterogeneity of biophysical the system. The observed spontaneous activity results from the way the connectivity interacts with the external inputs. We comment more on this point in section 3.8.3.

## 2.4 Methods

### 2.4.1 Recordings and experimental procedures

The database of spontaneous activity was collected by Liad Hollender and Péter Barthó in Rudgers University, U.S.A. The details of the surgeries and of the recordings presented here are retrieved from [123].

The data set is composed of two experiments of primary auditory cortex and eight experiments in primary somato-sensory cortex of Sprague-Dawley rats between 400 and 900 g. The recordings were performed with silicon probes (model Buzsaki 64-A64 , NeuroNexus, Ann Arbor MI) of 8 shanks; each shank was composed by 8 recording sites or "channels". The shanks were separated by  $200\mu m$ , (maximal separation between shanks: 1.4 mm). Contiguous recording sites had  $20\mu m$  of separation. The probe was inserted perpendicularly to the surface until attaining deep layers. Using histological reconstruction of the electrode tracks, electrode depth, and firing patterns, the precise recording location was estimated to be layer V. Rats were anaesthetized using urethane with a concentration of 1.5g/kg of body weight and ketamine (25-40 mg/kg). Urethane anaesthesia elicits periods of synchronized activity and desynchronized activity [25]. Depending on the synchronisation level, the recordings were segmented and aggregated into separate data sets of synchronised and desynchronised state.

The extracellular membrane potentials were high-pass filtered (1 Hz) and amplified (1,000 x) by using a 64-channel amplifier (Sensorium, Charlotte, VT), and digitized at 25 kHz (DataMax System, RC Electronics, Santa Barbara, CA) or 20 kHz (United Electronic Industries, Inc., Canton, MA).

### 2.4.2 Preprocessing the spike times into spike counts

We will go in over the method which with we bin the spike trains into spike counts. We use this method in most of our analysis, because it allows to put an upper-bound on the time scale of the measured correlations, but also because it offers a practical way of parsing and shuffling the data.

When dealing with electro-physiological data, can't do much with raw spike times, so we usually transform the signal composed of punctual times into a continuous signal, using a kernel, a function with a certain shape, that tends to zero in  $\pm\infty$ . Depending on the needs, one can use a Gaussian kernel (see figure 2.24), to better visualize the data, a exponential kernel to observe precise events), a Mexican hat to center locally the data (see Renart et al [123], S.I.) or a square hat, which will be our choice.

More precisely, we bin the spike times, with a given fixed time window  $t_c$  (which means that the square kernel is not centered on the spike time), counting how many spikes there are in the window (see figure 2.26). Every  $N_{jitt}$  bins (usually we take  $N_{jitt} = 5$  or 10), we make a second type of window, called a jitter window. Inside each jitter window, we compute the mean number of spike counts and we subtract it to all the bins inside the jitter window. Finally, if we want to assess correlation instead of covariance, we standardize the binned neuron by computing the standard deviation of this neuron and dividing the binned and centred counts of each cell by its standard deviation.

The rationale for such manipulation is based on the quantification of the covariances and correlations: as reminder, the covariance is the average value of the product of the deviations of two variables with their respective means:  $cov(x_1, x_2) = E[(x_1 - \mu_1)(x_2 - \mu_2)]$ . If the jitter window has the same length as the whole recording, then the mean  $\mu$  will be computed using the whole recording also. It might happen that one cell is dying along the recording and another fires more to compensate, so that, at the time scale of the whole recording, the covariance between both cells is negative. Even though, it could well be that in spite of the non-stationarity of both firing rates along the recording, both cells are positively correlated on a smaller time scale, and the only way to reveal this is to take a much smaller jitter window than in the previous example. In that way, in spite of the firing rate drifts, we can measure the joint co-fluctuation of both cells. The jitter window puts an upper bound on the time scale at which covariances are measured, and



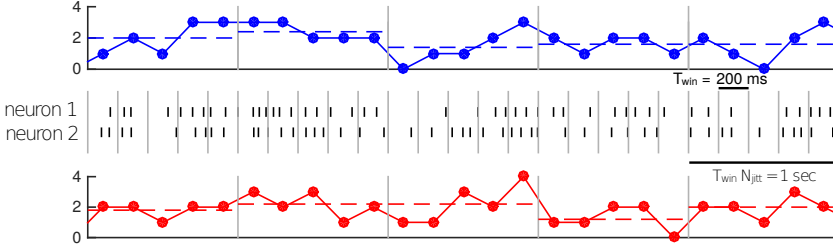


Figure 2.26: *Binning method that controls for non-stationarities, and that puts an upper bound on the time scale of the correlations. Starting from the raw spike times, for each spike train, we first count the spikes on windows of length  $t_c$  s, we remove the mean every  $N_{Jitter} \cdot t_c$  s, and finally we normalize by the standard deviation.*

isolates the correlations happening in one jitter window from another.

Thereby, we can create many partitions of the data. For example, if we number the jitter windows:  $1, 2, 3, 4, \dots, K$ , one can make series of  $W$  random permutations  $P_i$  of these jitter windows:  $P_1$ : (7, 40, 13, ... 23), ...,  $P_W$ : (61, 3, 90, ... 11). We can, for each permutation, split the data in halves and learn a model in the first half and test it on the second half, a process which we extensively use in this work.

Another advantage of such choice of preprocessing is that it offers a natural way of shuffling the data in a way that the variance of individual cells is preserved, but in which the correlations are destroyed: to obtain one shuffled version of the data, it suffices to randomly permute the spike counts inside the jitter windows. The measured correlations with such model constitute a null model that we can use to compare the data, and to estimate which principal components are significant (see chapter 4). This method is inspired from the jitter methods introduced by Amarasingham and Geman [3].

### 2.4.3 Coherence methods for determining the relative position of the shanks

For some of the recordings, we did not have the relative spatial location of the eight shanks. In order to recover this information, we analysed the

coherence of the raw extracellular signal. Then, using a simple algorithm, we found an order of the shanks in which the relative coherence of each pair of shanks decreases with the distance, which is what is to be expected.

The **coherence** between two real signals  $x(t)$  and  $y(t)$  at a given frequency  $f$ , is defined as:

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)}$$

$x(t)$  and  $y(t)$  might be for example the extracellular membrane potential of two cells (see figure 2.27 A). The coherence is a quantity between zero and one that defines the extent to which  $y$  results from  $x$  from a convolution: if  $y$  and  $x$  are such that  $y(t) = (h \star x)(t)$ ,  $h$  being a transfer function, the coherence between  $x$  and  $y$  is going to be equal to one.

$\hat{x}(f) = \int_{-\infty}^{+\infty} x(t).e^{-2\pi i f t}.dt$  is the Fourier transform of  $x(t)$ .

$G_{xx}(f) = |\hat{x}(f)|^2$  and  $G_{xy}(f) = \hat{R}_{xy}(t)$

$R_{xy}(t) = \int_{-\infty}^{+\infty} x(\tau)y(\tau+t).e^{-2\pi i f t}.d\tau$  is the cross-correlation of  $x(t)$  and  $y(t)$ .

Once we compute the coherence between two signals in a whole frequency band, we simply take the average coherence on this frequency band and compare its value with the value of other pairs.

The first step is to apply this analysis simultaneously to all the pairs of recording sites at high frequency (600 -1000 Hz) and to see which shank belongs each recording site (figure 2.27 D). We use such high frequency simply because the electric noise seems to be strong inside one shank and not coherent between shanks. We can see how the channels belonging to each shank cluster very well.

We then keep the indexes of the channels -excluding the dead channels- and average the lfp of all the channels belonging to each shank. Afterwards, for all possible pairs of shanks out of eight, we compute the coherence in a low frequency band between 6 and 20 Hz (figure 2.27 B,E).

We can see that the coherence doesn't decrease with the distance smoothly as it should. We then made a simple brute force algorithm in order to reorder the shanks.

### Algorithm:

- A is an 8 by 8 real matrix. The coefficients between the index (i,j) indicate the coherence between the shanks of initial indexes i and j. The objective of the algorithm is to find a new index order of the shanks.

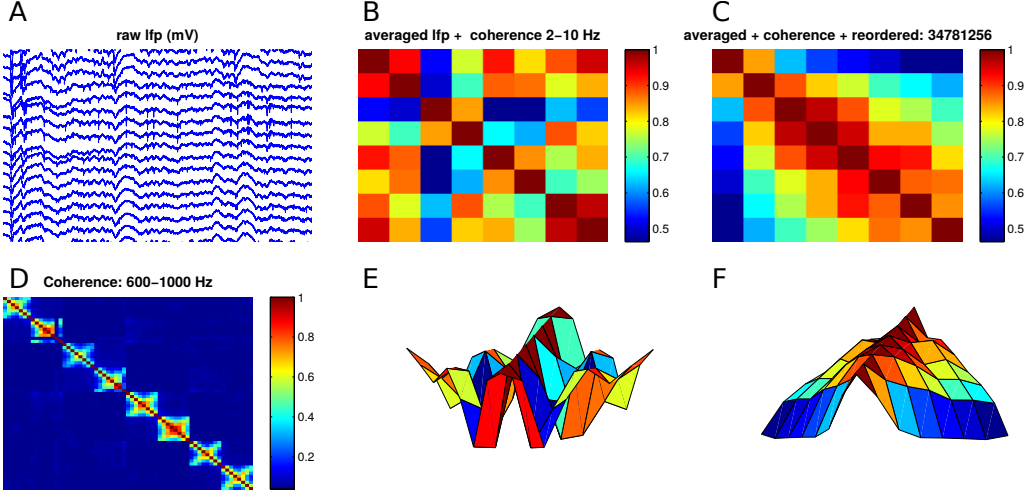


Figure 2.27: Reordering the shanks for experiment "Set 1". A: extracellular potential recorded at different channels. D: coherence at high frequency for all the channels. B: we average the signals of the channels within each shank and compute the coherence between the shanks. E : 3D view of the coherence matrix in B. C: reordered coherence matrix. F: 3D view of the reordered coherence matrix in C.

- D is an 8 by 8 overlap matrix with arbitrary coefficients, that has the shape we want A to have (The probe Buzsaki 64 has a linear arrangement of the shanks):

$$D = \begin{pmatrix} 0 & 10 & 4 & 3 & 1 & 0.5 & 0.2 & 0.01 \\ 10 & 0 & 10 & 4 & 3 & 1 & 0.5 & 0.2 \\ 4 & 10 & 0 & 10 & 4 & 3 & 1 & 0.5 \\ 3 & 4 & 10 & 0 & 10 & 4 & 3 & 1 \\ 1 & 3 & 4 & 10 & 0 & 10 & 4 & 3 \\ 0.5 & 1 & 3 & 4 & 10 & 0 & 10 & 4 \\ 0.2 & 0.5 & 1 & 3 & 4 & 10 & 0 & 10 \\ 0.01 & 0.2 & 0.5 & 1 & 3 & 4 & 10 & 0 \end{pmatrix}$$

- Generate and number all the permutations of the set  $\{1, 2, \dots, 8\}$  : ex  $1^{st}$  :  $p_1 = (1, 2, 3, 4, 5, 6, 7, 8)$ ,  $2^{nd}$ :  $p_2 = (2, 1, 3, 4, 5, 6, 7, 8)$ , ...
- for  $i = 1$  : Total number of permutations:

$$Cost_i = \sum_{j=1}^8 \sum_{k=1}^8 (A(p_i, p_i) \otimes D)_{(j,k)}$$

$\otimes$  is the point-wise matrix multiplication.

- Best order:

$$p_M : Cost_M \geq Cost_i, \forall i.$$

After the algorithm was applied, we can see how does the coherence decreases smoothly with the distance (figure 2.27 right top and bottom).

After applying this algorithm to all the experiments, we see that the order is pretty consistent among experiments that were done by the same person. In the following table 2.1 we see put the name of the experiment and the order found.

Experiment name:	Area:	Order:
setI Act 6650-7450 NoStim	A1	(3, 4, 7, 8, 1, 2, 5, 6)
setII Act 2200-2800 NoStim	A1	(3, 4, 7, 8, 1, 2, 5, 6)
c065a ActAll 0-320	S1	(8, 7, 6, 5, 4, 3, 1, 2)
c065b ActAll 0-540	S1	(8, 7, 6, 5, 4, 3, 1, 2)
c043 Act 500-1200	S1	(8, 7, 6, 5, 4, 3, 1, 2)
c048 ActAll 0-695	S1	(8, 7, 6, 5, 4, 3, 1, 2)
c057a ActAll 0-1090	S1	(6, 5, 7, 8, 2, 1, 3, 4)
c037 ActAll 0-575	S1	(4, 3, 2, 1, 5, 6, 8, 7)

Table 2.1: Summary table of experiments and shank order given by the algorithm

The only inconsistency between experiments resides in the last two experiments c057a and c037, in which the order found by the algorithm is not the same as the order found for the other experiments of somato-sensory cortex. Given that the algorithm finds perfectly consistent results for 6 out of 8 experiments, we proceeded to analyse the data with the corresponding orders.

#### 2.4.4 Methods on Spectral Analysis

We performed spectral analysis using the Chronux Matlab Package ([chronux.org](http://chronux.org)). In particular, we used the function `mtspectrumpt.m`, which uses a multitaper approach to calculate efficiently the power spectrum of a point process.

Briefly, multitaper techniques [113], [100] are designed to reduce the bias and inconsistency problems in the 'raw' estimate of the power spectrum (the square of the Fourier transform). Tapering (i.e., multiplying the data point-by-point by a suitable window) reduces the bias, and the use of several orthogonal tapers reduces the variance. The set of tapers given by the Slepian functions are optimal in that they are orthogonal and maximize the energy in a given time-frequency interval.

In order to calculate the power spectrum of a spike train from a given recording, we first rounded the time of each spike to 0.25 ms, and split the whole spike train into 5 s windows, in order to control for non-stationarities. The spectrum will be calculated for each window and the results will be averaged. We used all windows in which the neuron fired more than 2 action potentials. For each 5 s window, by manual inspection we found that a value for the time-bandwidth parameter  $TW = 5$  provided sufficient spectral smoothing without strong distortions. We used the recommended  $2TW - 1 = 9$  tapers to calculate the spectrum in each window.

The spectrum from each window was normalized by the mean power spectrum for all frequencies above a high frequency cut-off of 1000 Hz, which is equivalent to a normalization by the firing rate within than window (since the high frequency limit of the spectrum of a point process is the firing rate), and the final power spectrum for this neuron was equal to the average spectrum across all windows.

## 2.4.5 Non-parametric statistical tests

In this thesis, we make extensive use of non-parametrical statistical methods, in particular of permutation tests also known as exact tests. The strength of these tests is that they don't make assumptions about the underlying distributions. The first two following simple cases will give a clear overview of both the general framework of hypothesis testing, and of permutation tests.

### 2.4.5.1 Testing the significance of a linear fit

To test the statistical significance of an apparent linear relationship between two sets of observations  $\{x\}_{i=1,\dots,N}$  and  $\{y\}_{i=1,\dots,N}$ , we have to start by fitting a linear model by mean squares and obtain an ordinate and a slope.

In a second moment we create a certain number  $N_s$  of surrogate versions of the data. In each surrogate version, we apply a random permutation to the ordinates  $\{y\}$ , so that the correspondence between the  $\{x\}$  and the  $\{y\}$  is broken.

For each surrogate, we fit a linear model, and store the slope. The distribution of the slopes is our null model: if there is no statistical relationship between the  $\{x\}$  and the  $\{y\}$ , then the slope of the data will be situated inside the distribution. However, if the slope measured in the data has very low probability of having being generated by the null distribution, then we reject the null hypothesis. The p-value quantifies the proportion of samples from the null model that have a value that is as extreme as the measured value. We can compare the p-value to  $\alpha$ , the significance level and decide whether it is statistically significant.

We use this test several times, in particular to evaluate the significance of the relationship between the firing rate and the absolute value of the loadings, and also to test the statistical significance of the relationship between the mean power between 0-4 Hz of the spike trains and the loadings.

### 2.4.5.2 Statistical significance of the decrease of the standard deviation of the loadings with the distance

As referred in the main text, in order to test if the spread of the correlations decrease with the distance with respect to a local maximum, we first apply the following steps: for each experiment, we assign the loadings to the shanks they were recorded from using the method explained in a previous methods section (2.4.3). Then, for each shank, we compute the standard deviation of the loadings. Then, for each recording we find the location that has maximal standard deviation, and we normalize it, by dividing the standard deviation in each experiment by the maximal value of the standard deviation in this recording. Finally, we align the maximum across experiments.

If we then average naively across experiments, the mean standard deviation is going to decrease predictably with respect to the maximum, more or less independently of the spatial repartition of the loadings.

We then propose here a non- parametric statistical test to see whether the decrease in the data of the standard deviation of the loadings with respect to the maximum, is statistically significant compared with a null model in which the standard deviations of the loadings are distributed randomly in

space.

We create a series of surrogates: in each surrogate data set, the normalized loadings of all the experiments are put together and -omitting the maxima-, reassigned randomly to a different position.

We apply a parabolic least squares fit to the original data, and also to each surrogate version. In that way, we have a series of parabolas, that give the null model distribution of the absence of spatial decay with the distance. If the parabola fitted to the data is below the envelopes of the parabolas of the surrogates (or only above a small proportion of them), then we can assert that the decrease is statistically significant with respect to the null model, at this location. We can quantify this significance, for each location, using the p-value. The p-value at one location is the proportion of surrogate data, that take more extreme values than the parabolic fit of the data.

### 2.4.5.3 Reliability of the loadings

Projecting the network activity into the first principal component, gives the score, which is a time series of the same length as the spike counts. In order to evaluate the contribution of each cell to the score of the competitive activity, we compute the significance of the correlation between two time series: the spike count of one neuron and the score of the population computed excluding this neuron.

To know if this correlation is statistically significant, we compare it to the distribution of correlations of the score with many surrogate versions  $N_s$  of the cells activity  $\{t_i^s\}_{s=1,\dots,N_s}$ , in which we randomly permute the time counts inside the jitter window  $\tilde{n}_i(t)$ . Finally, we compute the p-value associated with each loading  $i$ , which is the number of surrogate statistics such that they are higher more extreme than the data statistic  $t_i$ . In summary:

$$\begin{aligned}
\forall i &= 1 \dots N \\
t_i &= \text{corr}(n_i(t), D_i.PC_1^i(t)) \\
\forall s &= 1 : N_s \\
t_i^s &= \text{corr}(\tilde{n}_i(t), D_i.PC_1^i(t)) \\
p_i &= \frac{\text{card}(\{t_i^s \mid |t_i^s| > t_i\})}{N_s}
\end{aligned}$$

## 2.4.6 Necessity and sufficiency of a given subspace of activity

### 2.4.6.1 Definitions and main computations

As a reminder, let's recall that when we have a multivariate data set  $D_1$  of dimension  $\langle T \rangle \cdot \langle N \rangle$ , of  $N$  variables, with  $T$  observations, in which all the variables are z-scored. We can project  $D_1$  into a vector  $\vec{v}$  of dimension  $\langle N \rangle \cdot \langle 1 \rangle$ , and obtain a time series of dimension  $\langle T \rangle \cdot \langle 1 \rangle$ . We can project back this time series onto the canonical basis and obtain a matrix of dimension  $\langle T \rangle \cdot \langle N \rangle$  and of rank one:  $D_1 \cdot \vec{v}_1 \cdot \vec{v}_1^t$ . Therefore, we can decompose the data  $D_1$  into the projection over a subspace  $\vec{v}_1$  and the projection along its orthogonal complement  $\{\vec{v}_i\}_{i=2, \dots, N}$ :

$$D_1 = D_1 \cdot \vec{v}_1 \cdot \vec{v}_1^t + \sum_{i=2}^N D_1 \cdot \vec{v}_i \cdot \vec{v}_i^t$$

We can then create a surrogate data set  $\tilde{D}_1$ , in which we destroy only the correlations over the desired subspace  $\vec{v}_1$ , to assess the necessity. We note  $\tilde{D}_1$  the shuffled version of a data set  $D_1$ . Because we use spike counts, by randomly permuting these spike counts inside the jitter windows, we preserve the variance of each neuron, but we destroy the covariance between the neurons.

$$D_1^* = \tilde{D}_1 \cdot \vec{v}_1 \cdot \vec{v}_1^t + \sum_{i=2}^N D_1 \cdot \vec{v}_i \cdot \vec{v}_i^t$$

Then we could compare the variance of the original data set  $D_1$  with  $D_1^*$ , but even better, we can compare the fraction of explained variance of  $D_1^*$  with another dataset  $D_2$  that was not used to generate  $D_1^*$ .

To do so, we compute the correlation matrices of  $D_1$ ,  $D_2$  and  $D_1^*$ ,  $C_{D_x} = \frac{1}{T} \cdot D_x^t \cdot D_x$ , and compare the explained fraction of variance among them.

The non-diagonal part of the correlation matrix,  $y_{(i,j)} = C_{(i,j)}^{D_2}$ ,  $j > i$  is taken as the predicted variable and we take the points of the triangular superior matrix of correlation of  $D_1$  or  $D_1^*$  as predictors,  $x_{(i,j)} = C_{(i,j)}^{D_1}$ ,  $j > i$  or  $x_{(i,j)} = C_{(i,j)}^{D_1^*}$ ,  $j > i$ . We choose a linear relationship between the predictors and the observed variables:

$$h_{(i,j)} = \alpha x_{(i,j)} + \beta$$



by least squares, we obtain  $\alpha$  and  $\beta$  such that  $\sum_{i,j} (y_{(i,j)} - h_{(i,j)})^2$  is minimal. Then, we obtain the fraction of explained variance  $R^2$ :

$$R^2 = 1 - \frac{(\sum_{k=1}^{N(N-1)/2} c_k - h_k)^2}{(\sum_{k=1}^{N(N-1)/2} c_k - \bar{c})^2}$$

For instance, in order to compute the fraction of explained variance by a particular subspace  $\vec{v}_1$  of a test set relative to the fraction of variance that the training set explains about the test set, we compute for each permutation the quantity:

$$\frac{R^2(D_1^*, D_2)}{R^2(D_1, D_2)}$$

And then we take the mean value.

#### 2.4.6.2 Algorithm

*D* : dataset, of dimensions  $< T > . < N >$

$D \leftarrow \text{preprocess}(D)$  (binning, centering with fixed jitter windows,  $z$ -score)

Split  $D$  in two :  $D_1$  (training),  $D_2$  (test set)

Choose  $B = [\vec{v}_1, \dots, \vec{v}_P]$ , an orthogonal subspace to test

Deduce  $B^\perp = I - BB'$ , the orthogonal complement of  $B$

$\forall n \in [1, N_{\text{surrogates}}]$

$$D_1^\diamond = D_1.B.B' + \tilde{D}_1.B^\perp(B^\perp)', \quad C_1^\diamond = \frac{1}{T} \cdot (D_1^\diamond)^t \cdot D_1^\diamond$$

$$D_1^* = \tilde{D}_1.B.B' + D_1.B^\perp(B^\perp)', \quad C_1^* = \frac{1}{T} \cdot (D_1^*)^t \cdot D_1^*$$

$$\frac{R^2(D_1^*, D_2)}{R^2(D_1, D_2)} \quad \frac{R^2(D_1^\diamond, D_2)}{R^2(D_1, D_2)}$$

Index of necessity :  $\text{mean}(\frac{R^2(D_1^*, D_2)}{R^2(D_1, D_2)})_n \in [0, 1]$

Index of sufficiency :  $\text{mean}(\frac{R^2(D_1^\diamond, D_2)}{R^2(D_1, D_2)})_n \in [0, 1]$

If by destroying the correlations over the chosen subspace  $\{\vec{v}_1, \dots, \vec{v}_P\}$  the fraction of explained variance or "Index of necessity" is small, it means that the chosen subspace is necessary to explain for the correlation structure of the data.

If by preserving the correlations over the chosen subspace  $\{\vec{v}_1, \dots, \vec{v}_P\}$  the fraction of explained variance or "Index of sufficiency" is big, it means that the chosen subspace is sufficient to explain for the correlation structure of the data. This method is summarized graphically in the figure [2.28](#).

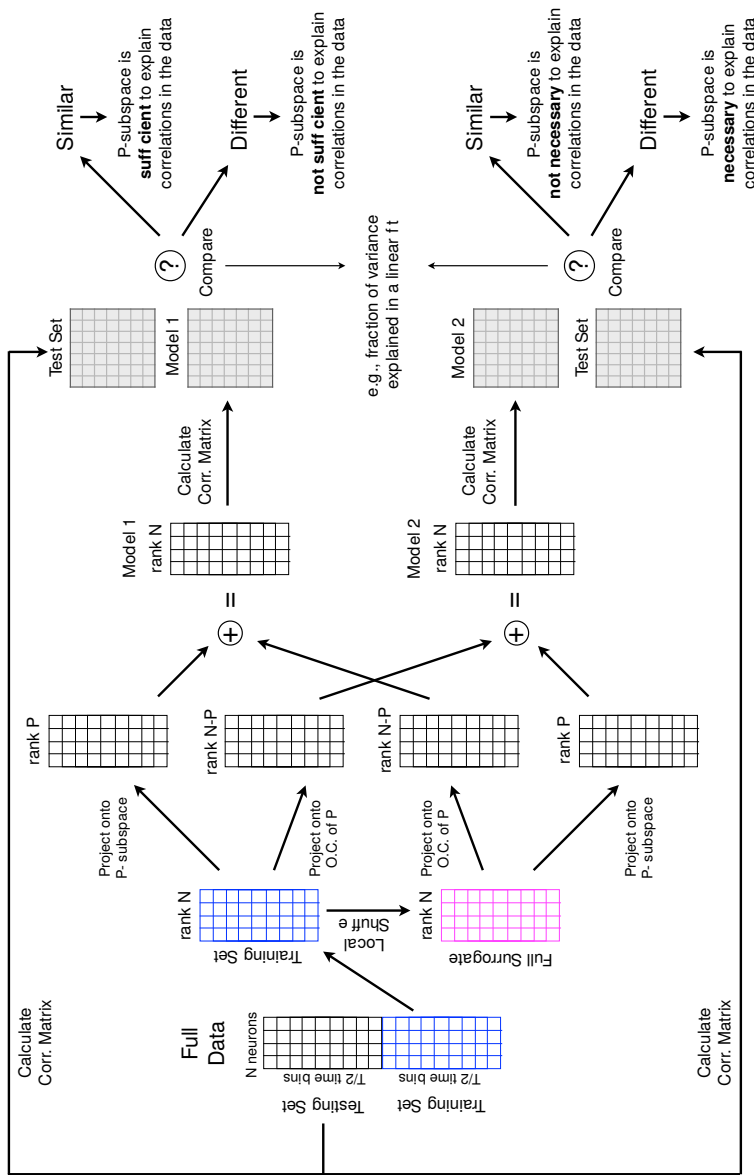


Figure 2.28: Method to assess necessity and sufficiency of a given subspace in order to explain the variance in a data set.



# Chapter 3

## Modelling of competitive activity

### HIGHLIGHTS

- In this chapter, we explore patterns of connectivity and mechanisms that might generate competitive amplification.
- We first study the balanced randomly connected excitatory and inhibitory network, and show that in this network there is not a pattern of activity that is more amplified with respect to the others.
- We propose two simple models (NCA and TCA) generating competitive amplification through different mechanisms.
- One of the two models - TCA- makes robust predictions that match the experimental observations.
- We extend the low dimensional TCA model to a high dimensional version, and make additional predictions that are in agreement with the data.
- We also study other candidate circuits that generate competition.

## Contents

---

<b>3.1</b>	<b>Introduction: dynamics in neural circuits . . . .</b>	<b>96</b>
3.1.1	Motivation . . . . .	96
3.1.2	Modelling choices . . . . .	96
3.1.3	Amplification in computational neuroscience . .	99
3.1.4	Different mechanisms generating amplification in neural networks . . . . .	100
<b>3.2</b>	<b>Study of the E-I randomly connected network .</b>	<b>106</b>
3.2.1	Network architecture and simulation . . . . .	106
3.2.2	The eigenspectrum of the correlation matrix doesn't reveal competitive activity . . . . .	107
<b>3.3</b>	<b>Circuits generating competitive activity . . . . .</b>	<b>109</b>
3.3.1	The two dimensional linear excitatory and inhibitory network . . . . .	109
3.3.2	Normal competitive amplification (NCA) . . . .	112
3.3.3	Transient competitive amplification (TCA) . . .	114
<b>3.4</b>	<b>Model predictions and comparison with the data</b>	<b>117</b>
3.4.1	NCA and TCA generate negative correlations among $E_1$ and $E_2$ . . . . .	117
3.4.2	Asymmetry in the connectivity generate differ- ence in the variances of $E_1$ and of $E_2$ . . . . .	118
3.4.3	Asymmetry in the connectivity generate differ- ence in covariances between the two populations and the inhibitory neurons . . . . .	118
3.4.4	NCA crosscorrelogram is symmetric while TCA has a lag in the crosscorrelogram . . . . .	119
3.4.5	Angle between the PC1 and the uniform . . . . .	119
3.4.6	Comparison with data . . . . .	120
<b>3.5</b>	<b>Generating competitive low dimensional dynam- ics in a high dimensional network . . . . .</b>	<b>123</b>
3.5.1	Limitations of the low dimensional motif . . . . .	123

3.5.2	Generating an expanded high dimensional low-rank matrix . . . . .	124
3.5.3	Adding noise to the connectivity . . . . .	126
<b>3.6</b>	<b>Predictions of the high dimensional model . . .</b>	<b>130</b>
3.6.1	Correspondance between the HD and the LD version of the model . . . . .	130
3.6.2	Comparison between data and model . . . . .	130
<b>3.7</b>	<b>Other models of competition . . . . .</b>	<b>133</b>
3.7.1	Alternative low dimensional motif to TCA . . . .	133
3.7.2	Translation-invariant connectivity matrices might generate competitive activity, but with different characteristics . . . . .	135
<b>3.8</b>	<b>Summary and discussion . . . . .</b>	<b>138</b>
3.8.1	Summary . . . . .	138
3.8.2	Relation with previous work . . . . .	139
3.8.3	Functional relevance of the asymmetry . . . . .	141
3.8.4	On the specificity of model predictions . . . . .	142
3.8.5	Model fitting . . . . .	142
3.8.6	More experiments to refine the predictions ? . . .	144
3.8.7	Other possible experiments to expand our understanding about the role of competition . . . . .	144
<b>3.9</b>	<b>Methods . . . . .</b>	<b>146</b>
3.9.1	Balanced randomly connected network simulation	146
3.9.2	The Lyapunov equation . . . . .	146
3.9.3	Mean trajectory and cross-correlograms for linear dynamical systems . . . . .	147
3.9.4	Cross-correlograms of spike trains . . . . .	147
<b>3.10</b>	<b>Appendix . . . . .</b>	<b>149</b>
3.10.1	Two dimensional linear excitatory and inhibitory network . . . . .	149
3.10.2	Study of the correlation in the 2D linear network	150

3.10.3 Rotating the covariance matrix from the Schur to the canonical basis . . . . .	152
3.10.4 Spectral decomposition . . . . .	153
3.10.5 Computing the eigenvectors of a connectivity matrix . . . . .	154
3.10.6 Schur decomposition of a 2*2 matrix . . . . .	156
3.10.7 Adding the deterministic and the stochastic high dimensional connectivity matrices . . . . .	157
3.10.8 Summary of the implementation of the high dimensional model . . . . .	163
3.10.9 Translation-invariant connectivity matrices . . .	167

---

## 3.1 Introduction: dynamics in neural circuits

### 3.1.1 Motivation

Having observed the competitive dynamics in deep layers of auditory and somato-sensory cortex, we would like to have a mechanistic understanding of this phenomenon. In particular, we would like to know which kind of connectivity matrix might generate these dynamics, but even more, how can we understand functionally how different connectivities lead to different dynamics. We will study different models that lead to different dynamical observables that we will contrast with the data. Another important question that we will answer is, under which conditions on a high dimensional connectivity matrix, can low-dimensional activity be generated.

In this introductory section, we will start by explicating the modelling choices, the most important of which is the computational framework of amplification.

### 3.1.2 Modelling choices

The brain is capable of doing very complex operations like learning the structure of the world and act on it. All these different operations -sensory processing, memory, language, decision making and motor activities- are implemented though the joint actions of ensembles of neurons.



The process of modelling and understanding a neural network involves making assumptions and combine mathematical objects to obtain answers: which input is the network receiving from the outside, what are the operations that single neurons perform and finally, how are these neurons connected. Models including many biophysical details usually have a high number of parameters, which makes them very flexible. However, these kind of complex models might be also harder to understand. The challenge of modelling these circuits is to postulate which elements are fundamental to focus on the few biophysical elements that we think are capital for the phenomenon we want to reproduce. Simpler models are then easier to falsify, because they state clearly which are the key factors that cause a phenomenon.

As said by Hansel and Sompolinsky, "simplified abstract models offer very valuable theoretical tools to gain insight into the working of these systems. Not only is the reduced parameter space of these simplified models significantly easier to search, but many are amenable to analytical investigations. Analytical solutions are extremely useful in that they often explicitly reveal the important relationships between a dynamic property of the network and some of its parameters" [83].

#### 3.1.2.1 Rate models and spontaneous activity

All around this chapter, we will use the formalism of rate models, in which the neuron-like units output directly firing rates, rather than action potentials. More precisely, what we will usually call the rate  $x_i(t)$  represent in fact the deviation in firing with respect to the mean firing rate of that cell:

$$x_i(t) = \tilde{x}_i(t) - \bar{x}_i$$

$\tilde{x}_i(t)$  represents the true instantaneous firing rate of cell  $i$ , and  $\bar{x}_i$  its average over time. This quantity has a simple operational interpretation when we consider evoked activity: we present a stimulus several times, and each time we record a spike train  $\rho(t) = \sum_i \delta(t - t_i)$ . Across trials, we can bin, average and smooth the number of spikes by unit of time, and obtain a firing probability density  $\langle \rho(t) \rangle$ . The true instantaneous firing rate is:

$$\tilde{x}_i(t) = \frac{1}{dt} \int_t^{t+dt} \langle \rho(\tau) \rangle .d\tau$$

When we are dealing with spontaneous activity, the notion of firing probability density is less clearly defined, because we don't have a notion of trial,

so experimentally we can't have access to it. When we analyse the data, we make an assumption about the time scale at which the neurons sum their inputs, by putting a kernel of a certain width on top of each spike (see methods 2.4.2). We then make the assumption that the smoothed version of the spike train is the firing probability density and the firing rate. Biophysically, slow membrane dynamics could be causing some temporal smoothing on the spikes, but whether the relevant quantity for neurons are the precise spike times or rather the firing rates, is a major question that is still debated.

By using rate models instead of deterministic models of spiking networks, we are being conservative in the level of detail we are modelling, because we are making claims about a coarser quantity, and not about the precise spike times. Even if we had a very precise spiking model of the network, we could not make those predictions because we don't have experimental access to the input currents [33]. Also, we use these rate models for practical reasons, because they are easier to simulate than spiking networks, and offer more analytical power because they use continuous variables. Rate models have been used with great success for modelling neural systems. However, it is known that these models are not perfect, notably that they fail in capturing spiking effects at small time scales and in cases in which neurons show important synchronous firing [33].

#### 3.1.2.2 Modelling spontaneous activity using stochastic linear dynamical systems

As Murphy and Miller [105] and Hennequin et al. [64], we use the mathematical framework of linear dynamical systems to try to reproduce the competitive dynamics observed during desynchronised states. We consider a recurrent network of  $N$  neurons whose dynamics are governed by the joint differential equation (in matrix notation):

$$\dot{X} = -X + JX + B\xi$$

$X$  is a  $N$  by 1 vector of firing rates. The derivative of the firing rate  $\dot{X}$  results from the superposed effects of three factors: first, the decay of the system ( $-X$ ) towards the steady state  $\dot{X} = X = \vec{0}$ . Second, the post-synaptic currents resulting from the activity  $X$  on the recurrent connectivity  $J$ . Third, the noisy post-synaptic currents from outside the network:  $B$  is a connectivity matrix that might correlate the inputs into cells. Un-

less otherwise stated, we will consider  $B$  to be an identity matrix, so that neurons receive uncorrelated inputs with unitary variance.  $\xi$  is a vector of uncorrelated Gaussian processes of zero mean and unit variance: each process represents a large sum of stochastic variables, like currents from contiguous circuits.

This simple differential equation summarizes well the modelling choices and the assumptions about spontaneous activity, already mentioned in figure 1.5: the dominant patterns of observed activity result from the interaction of the recurrent connectivity  $J$  with uncorrelated external inputs  $\xi$ . Even more precisely, we will explain how this computational framework, predicts a profound link between some features of the connectivity and the direction in neural space of the most observed activity patterns.

This differential equation with a stochastic term is called an **Ornstein-Uhlenbeck process**. The aim of this model is not to reproduce particular firing sequences, that can be seen in the data, but rather to describe statistically the mean response of the system. For example in an autapse if we plug a constant input  $I$  many times, like in figure 3.1, and we inject external noise  $\xi$ , the responses across trials will be very different. However, if we compute the average response across trials, we will see that the autapse on average increases its firing rate and converges to a steady state, like the voltage of a capacitor in a RC circuit. Finally, biggest advantage of using the formalism of stochastic process, is that it is simple and it allows us to understand analytically the magnitude and the temporal structure of covariances in a recurrent network.

Following Murphy and Miller [105], the main assumption that we will make in this chapter is that during the active state, local circuits receive unstructured noisy inputs from the outside and then the recurrent connectivity shapes the dynamics, selectively amplifying certain patterns of activity.

#### 3.1.3 Amplification in computational neuroscience

Before explaining in detail the mechanisms by which certain activity patterns are enhanced with respect to others through what is called amplification, let's briefly mention that many problems in computational neuroscience can be studied with similar mathematical tools and ideas, so they could be cast as examples of "amplification". Amplification has brought an answer to how persistent activity in the brain is generated by neurons with relatively

short time constants [6], [20], [52]. Also, amplification has been used in the context of understanding how to select a stimulus, and modulate it multiplicatively [57] and also in the context of perceptual decision-making tasks, using winner-take-all networks [155], [160], [129], [128]. Amplification has also been useful in the context of feature selectivity to explain orientation tuning or contrast invariance [11], [83]. Finally, amplification has been fruitful to model why spontaneous and evoked to oriented stimuli activity of visual cortex seem so similar [80], [147],[105].

### 3.1.4 Different mechanisms generating amplification in neural networks

In this paragraph, we will recall briefly two mechanisms that have been described in the literature to amplify particular directions of activity.

#### 3.1.4.1 Normal amplification

We will start by the simple scalar case of a single autapse - a neuron connected to itself- to which we plug a step of current  $I$ .

$$\dot{x} = -x + Jx + I + \xi$$

By removing the stochastic noise  $\xi$ , we solve the equation and obtain the average firing rate, taking  $x_{(t=0)} = 0$  as initial condition.

$$x_{(t)} = \frac{I}{1-J}(1 - e^{-\frac{t}{\tau}})$$

The more the autapse  $J$  is close to 1, the bigger the steady state value of the firing rate, but also the more time the system takes to reach the steady state: the time constant is  $\tau = \frac{1}{1-J}$  (see figure 3.1).

Lets now consider the high dimensional case of a recurrently connected network which connectivity is normal, e.g symmetric. As the connectivity matrix  $J$  is diagonalizable for normal matrices, it can be expressed as a product of a rotation matrix  $V$  and of a diagonal matrix  $D$  whose diagonal elements are the eigenvalues  $\lambda$  of  $J$ :  $J = VDV^{-1}$ . Rotating the system with the rotation matrix  $V$ , the new axis -which are the columns of  $V$ - are the eigenvectors of  $J$ . In this new basis, each variable behaves like the autapse

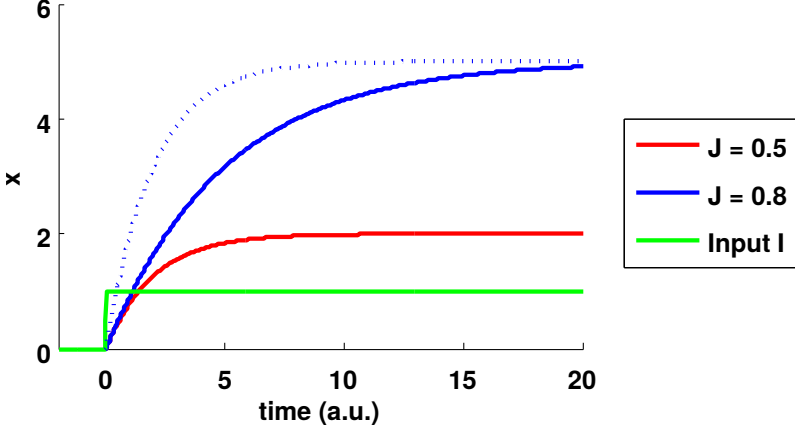


Figure 3.1: Example of normal amplification, for two values of  $J$ , small in red and strong in blue. The dashed line is the scaled version of the red one, to show the difference in convergence time to the asymptote. In green the input current. Inspired from [105].

we studied previously, and the eigenvalues  $\{\lambda_i\}_{i=1,\dots,N}$  are the strength of each of these autapses. We can rotate the system into this new basis:

$$\begin{aligned}
 (E) : \quad & \dot{X} = -X + JX + I + \xi \\
 (E) \iff & \dot{X} = -X + VDV^{-1}X + I + \xi \\
 \iff & V^{-1}.\dot{X} = -V^{-1}.X + DV^{-1}X + V^{-1}.I + V^{-1}.\xi \\
 \iff & \dot{Y} = -Y + D.Y + \tilde{I} + \tilde{\xi}, \quad Y = V^{-1}.X, \quad \tilde{I} = V^{-1}.I, \quad \tilde{\xi} = V^{-1}.\xi \\
 \iff & \dot{y}_i = -y_i + \lambda_i.y_i + \tilde{I}_i + \tilde{\xi}_i, \quad i = 1, \dots, N
 \end{aligned}$$

As we consider that the effective external noise is uncorrelated  $B = I$ , in the new basis, the noise  $\tilde{\xi}$  is also uncorrelated across cells.

One interesting dynamical interpretation of the high dimensional system, is that if -in absence of noise - we initialize the system with  $X_{(t=0)}$  along the eigenvector  $\vec{v}$ , the derivative of the direction  $\dot{X}$ , has the same direction as  $\vec{v}$ , and then the system will remain in the same direction over time  $\dot{\vec{v}} = (J - I)\vec{v} = (\lambda_v - 1)\vec{v}$ , until it is totally damped  $X(t) \approx \alpha_{(t)}.\vec{v}$ , with  $\alpha_{(t)} \rightarrow 0$  when  $t \rightarrow +\infty$ .

The noise plugs at each time point new directions that can be amplified, but the majority of directions are quickly quenched, exempt the ones that have longer time constants. The **negative feedback** measures how quick the system is pushed back towards the equilibrium in a particular direction: it is quantified as the real part of the eigenvalue minus the decay of the system ( $\Re(\lambda_i) - 1$ ). The system is stable if the real part of the maximum of these eigenvalues is smaller than the unitary decay of the system.

The negative feedback of an eigenvector determine the associated time constant:  $\tau_i = \frac{1}{1-\Re(\lambda_i)}$ . When an eigenvalue is equal to the decay  $\lambda_i = 1$ , the system behaves like a perfect integrator and then all the noise that arrives is accumulated, which makes the firing rate increase until the system explodes. When one or more of the eigenvalues of the system are close to the instability, inputs in the direction of these eigenvalues will be integrated by the network, and variance along these dimensions will grow. We refer to this phenomenon as "normal amplification", because it is the only way in which the recurrent dynamics of a normal network can amplify its input.

The definition of the amplification during spontaneous activity that we proposed in the introduction:  $X_{(t)} \approx \alpha_{(t)} \vec{v}$  is consistent with these considerations, and postulates a strong link between the observed patterns of activity and an idiosyncratic characteristic of the underlying connectivity -the eigenvectors of J-.

Normal amplification is the traditional way by which amplification has been thought to be generated in cortical circuits [38]. It is important to emphasize that there is an inherent trade-off between the magnitude and the time speed of fluctuations amplified by this mechanism. With this mechanism, we can only produce large fluctuations that are very slow. In addition, because amplification requires that one or more eigenvalues are at the edge of the stable region, there is also a trade-off between normal amplification and dynamical-stability.

#### 3.1.4.2 Non-normal amplification

Non-normal amplification was introduced initially by Ganguli et al. [45], and then by Murphy and Miller [105] and by Goldman [52] in 2009, as an alternative way of generating amplification. The non-normal amplification was a known mechanism in the domain of fluid mechanics, but it had not been applied to computational neuroscience before.

Considering a linear dynamical system of size  $N$ , we can show, that for a very broad class of real connectivity matrices  $J$  called non-normal matrices (for which  $JJ^t \neq J^tJ$ ),  $J$  can be decomposed as:  $J = P.T.P^t$ , where  $P$  is a unitary matrix and  $T$  is a upper-triangular complex matrix, whose diagonal elements are the eigenvalues of  $J$ . This means that a system with a recurrent connectivity matrix (which is non-normal) can be rotated into a system which has a feed-forward connectivity, on top of loopiness, summarized along the diagonal (eigenvalues). Feed-forward connectivity and upper-triangular matrices mean the same thing.

$$\dot{X} = -X + JX + I + \xi \iff \dot{Y} = -Y + TY + \tilde{I} + \tilde{\xi}, \quad Y = P^t.X$$

For example, lets place ourselves directly in the rotated space: consider a two dimensional system, a feed-forward connectivity matrix  $T = \begin{pmatrix} \lambda_1 & g \\ 0 & \lambda_2 \end{pmatrix}$ .

Given the initial conditions:  $\begin{pmatrix} x_1(t=0) \\ x_2(t=0) \end{pmatrix}$ , the differential system is:

$$\dot{X} = \begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = - \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} \lambda_1 & g \\ 0 & \lambda_2 \end{pmatrix} \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$$

Solving the system we obtain (for  $\lambda_2 \neq \lambda_1$ ):

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} x_2(0) \frac{g}{\lambda_2 - \lambda_1} \cdot (e^{(\lambda_2 - 1)t} - e^{(\lambda_1 - 1)t}) + x_1(0)e^{(\lambda_1 - 1)t} \\ x_2(0) \cdot e^{(\lambda_2 - 1)t} \end{pmatrix}$$

We see in figure 3.2 that in such a feed-forward network, we can have a transient amplification of  $x_1(t)$ , because the second neuron injects current in the first neuron. The bigger  $g$ , the bigger the transient is going to be. It is then clear that to see such kind of non-contractive dynamics, we need at least two units (we can call them neurons or populations of neurons). One autapse, can not generate such transient amplification. In systems with connectivity matrices in which  $J.J^t = J^t.J$ , called **normal matrices**, we can not observe such non-contractive dynamics [64]. With those matrices, we can only observe the kind of amplification described in the last paragraph.

The key element in understanding how to rotate from the **canonical basis** -the neural basis-, into the Schur basis, resides in the change of basis matrix  $P$ :  $P^t.J.P = T$ . This matrix is built by orthogonalizing the

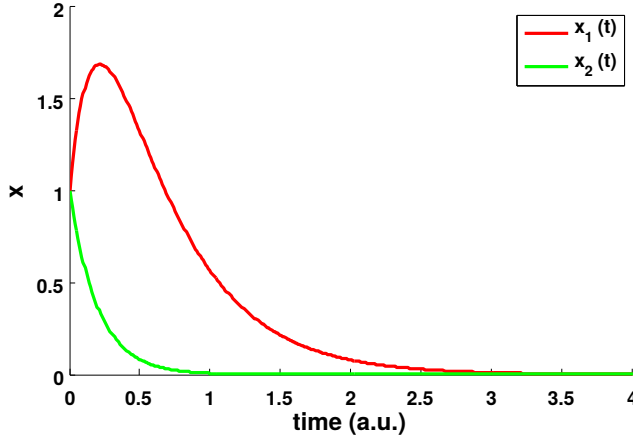


Figure 3.2: Example of contractive  $-x_2(t)-$  and non-contractive dynamics  $-x_1(t)-$ . We choose  $(x_1(0) = 1, x_2(0) = 1, \lambda_1 = -1, \lambda_2 = -4, g = 10, I = 0, \xi = 0)$ . Inspired from [105].

eigenvectors  $\{v_i\}_{i=1,\dots,N}$  of  $J$  through the Gram-Schmidt orthogonalisation procedure:

$$\begin{aligned} u_1 &\leftarrow v_1 \\ u_2 &\leftarrow v_2 - u_1(v'_2.u_1) \\ u_3 &\leftarrow v_3 - u_1(v'_3.u_1) - u_2(v'_3.u_2) \\ &\dots \end{aligned}$$

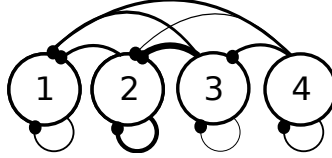
The new basis  $\{u_i\}_{i=1,\dots,N}$  is orthonormal and it is called the **Schur basis**. This basis is not unique because it depends on the order in which choose the vectors in the Gram-Schmidt procedure. The columns of  $P$  are formed by the vectors that compose this Schur basis.

A property of normal matrices, is that its eigenvectors are orthogonal ( $v'_j.v_i = 0, \forall i, j$ ), so the eigenbasis is equal to the Schur basis, and the non-diagonal elements of  $T$  are nil, so that  $T = D$ .

The Schur representation of the connectivity ( $J$ ), or equivalently of the dynamical system  $(J - I)$  admits a schematic representation that we will use thoroughly in this work. In figure 3.3, we choose a four dimensional



system. Each unit may receive inputs from units that have a number equal (recurrent connection) or bigger than the number of the unit in the chain. We might use the notation  $t_{ij}$ , to denote the strength of the link from unit  $j$  to unit  $i$ .  $j \geq i$  because the matrix  $T$  is upper-triangular.



*Figure 3.3: Schematic representation of a Schur decomposition. The links between units represent the non-diagonal elements of the upper-triangular matrix. The recurrent links represent the negative feedback. The width of the links represents the strength of the connection. The number  $i$  of the amplified patterns  $u_i$ , also called modes, is the same as the one obtained during the Gram-Schmidt decomposition.*

The most important aspect about the Schur decomposition, is that with this simple rotation, we are doing a change of axis, from the canonical basis to the Schur basis. The new axis point in the directions given by the Schur basis  $\{u_i\}_{i=1,\dots,N}$ . Moreover, these directions (also known as **patterns** or **modes**), because they are linear combinations of the neurons, can have a functional interpretation, as shown by Murphy and Miller [105]: for example as we will see further on, when the variance along the uniform direction

$\vec{u} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix}$  is very strong with respect to the variance of the other modes, all neurons go up and down together.

The Schur representation of the connectivity gives a very insightful way of qualitatively understanding the dynamics, because we have an explicit representation of the amplified patterns, and the way they are linked to each others. Because this is an orthogonal change of basis, it preserves the total activity variance. In complement of the Schur representation, we must do other computations to assess quantitatively how much a specific set of parameters amplify certain modes rather than others. For example, we can simulate the network, compute the covariance matrix and also do PCA on the covariance matrix.

## 3.2 Study of the E-I randomly connected network

As mentioned in the introduction, the randomly connected network, is the standard model of desynchronised activity. Van Vreeswijk and Sompolinsky [149] showed that due to the balanced state, the randomly connected network reproduces the histogram of firing rates. Renart et al. [123], showed that in spite of the fact that the external population of neurons inject correlated input into the two other populations, the balanced network (strong and densely connected) settles down into a steady state in which neurons have a near zero, slightly positive, mean pairwise correlation, like observed during the desynchronised state.

### 3.2.1 Network architecture and simulation

The randomly connected network is composed by one external population of neurons that project in a feed-forward manner to two recurrently connected groups of excitatory and inhibitory neurons figure 3.4 A.

Using Brian, a spiking network simulator [53], we simulated the balanced recurrent network, using the same parameters than in the original Renart 2010 paper [123] (see methods 3.9.1). In figure 3.4 B we see the raster plot of 5s of activity.

In figure 3.4, we replied the model predictions, plotting the histogram of pairwise correlations (C), and the histogram of firing rates (D), which is unimodal and skewed, like in [149], [123]. The correlations where computed binning the spike trains into time windows of  $t_c = 0.1$  s and taking a jitter window of ten time windows.

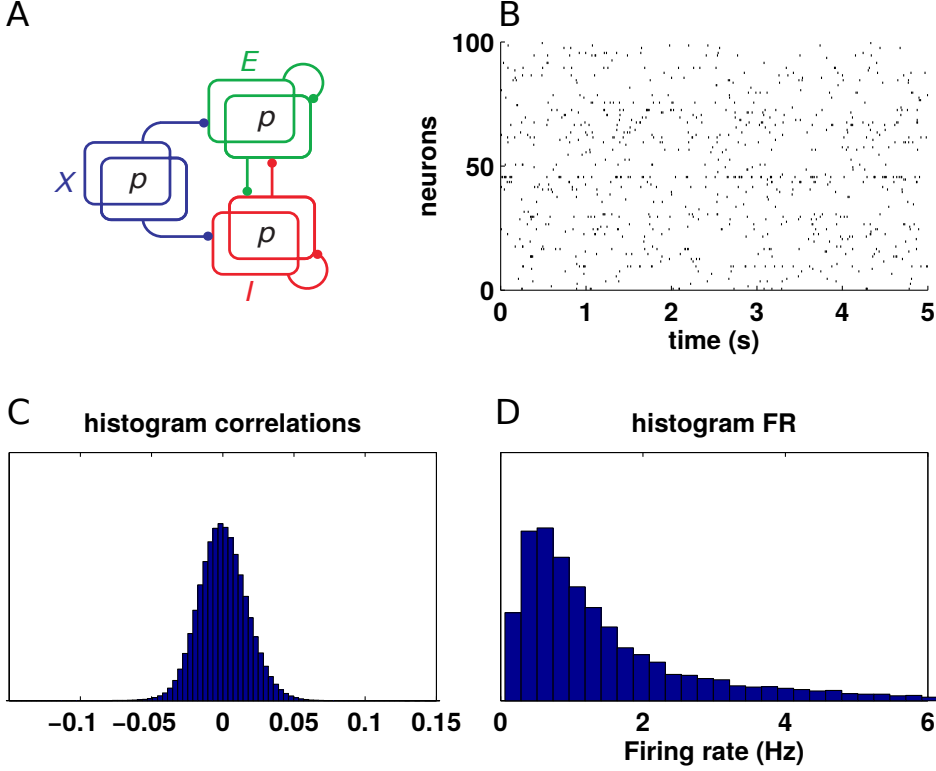
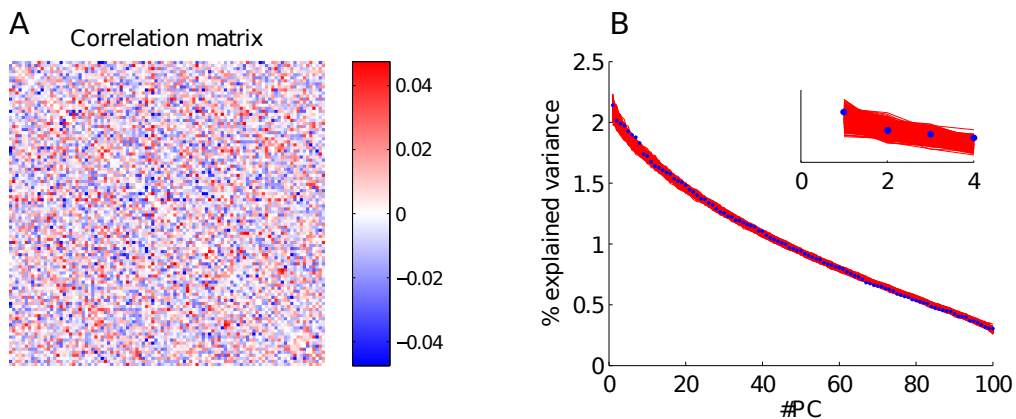


Figure 3.4: Randomly connected balanced network as described in [123]. A: Network architecture: one external excitatory population plugs feed-forward shared input into the network. B: raster plot of 100 randomly picked neurons for five seconds. C: histogram of pairwise correlations. D: histogram of firing rates. Parameters:  $t_c = 0.1$  and  $N_{jitt} = 10s$ .

### 3.2.2 The eigenspectrum of the correlation matrix doesn't reveal competitive activity

Randomly sampling 100 neurons, out of the 5000 simulated, we computed their correlation matrix, and its associated eigenspectrum. We compared the eigenspectrum of the data with 1000 eigenspectra of surrogate versions of the data, in which the correlations were destroyed, by permuting the spike counts inside each jitter window, maintaining the variance, and destroying the covariance.



*Figure 3.5: Randomly connected networks doesn't display competitive activity. A: correlation matrix, in which the diagonal was removed. B: eigenspectrum (in blue), and associated eigenspectra of the data shuffles (red). For visibility reasons, we linked the eigenvalues of each surrogate by a red line. Number of surrogates: 1000. Parameters:  $t_c = 0.1$  and  $N_{jitt} = 10$ .*

As we see in figure 3.5 B, the eigenspectrum falls into the envelope formed by the distribution of surrogate eigenspectra (see chapter 5). The envelope of surrogate eigenspectra was obtained by computing many times the eigenspectrum of surrogate data. This result is coherent with a recent paper [158], which precisely simulates the balanced network on a spiking network, does factor analysis (a technique closely related to PCA, see section 5.1) and finds a flat eigenspectrum.

The main point of this analysis is that in the balanced network, activity is not temporally structured -at least in a way that can be detected using PCA-, whereas in the recordings we saw that it has temporal structure. The fact that the mean pairwise correlations is close to zero - with approximately 50 % of the pairs positively and 50 % of the pairs negatively correlated-, is not enough to imply competitive dynamics. In the randomly connected balanced network, the tails of the correlation histogram are small whereas in the data, they appear to be large and approximately cancel. Finally, in the balanced network, there is not a pattern of activity that is significantly more amplified than others, as we see in the data (see figure 2.3).

We conclude from this analysis that the E-I network with random connectivity is not enough to generate competition, and then we need to add

more structure to the connectivity matrix, in order to achieve competition.

### 3.3 Circuits generating competitive activity

With the aim of adding structure to the connectivity matrix we could follow two strategies:

The first would be to use a systems identification approach, using the data to infer both the connectivity matrix  $J - I$  - the "drift" in physical jargon- and the noise matrix  $B$  - the "diffusion". Racca and Porporato [121] propose one methodology for this identification in the scalar case.

The second possibility -which is the one we followed- is to identify circuits that might generate competitive activity. Our approach will be, for the first part of this chapter, to focus on low dimensional models, as simple as possible, in an effort to identify possible mechanisms generating the kind of competitive dynamics we see in the data. Later on we will use this knowledge to build high-dimensional networks in order to address the question of how the low-dimensional correlation structure that we observe may come about.

We will start by introducing the two dimensional linear excitatory and inhibitory network, treated by Murphy and Miller [105]. This network will allow us to present the parametrisation of the network, and more importantly, it will be a starting point of subsequent 3D models. Each of the models we will present will be a bit more complex than the previous ones, maintaining some features of the simpler models.

#### 3.3.1 The two dimensional linear excitatory and inhibitory network

##### 3.3.1.1 Presentation of the network

In the appendix section 3.10.1, we present the linear network with two populations of neurons in a similar way as Murphy and Miller [105] did in their paper. We model the temporal evolution of the variables  $x_E$  and  $x_I$ , which could represent the mean firing rate of a population of excitatory and inhibitory neurons. For the moment, we will consider take these models as a face value, in the sense that we are not thinking of them as low dimensional representations of high dimensional networks. In that sense, these models doesn't contain the network size  $N$ .

We can parametrise this network as follows:

$$J = \begin{pmatrix} J_E & -\Delta - J_E \\ J_I & -\Delta - J_I \end{pmatrix}$$

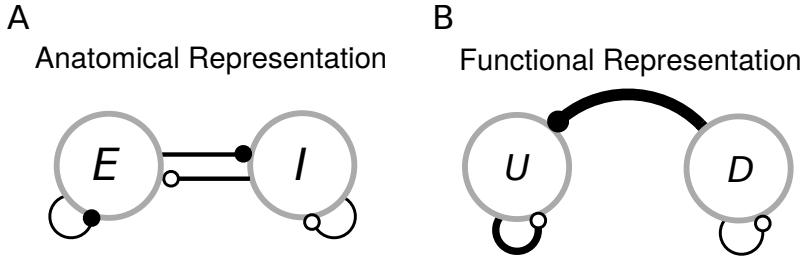


Figure 3.6: Anatomical (A) and functional representation (B) of a network representing two populations of excitatory and inhibitory neurons, as shown by Murphy and Miller [105].

All the parameters ( $J_E, J_I, \Delta$ ) are positive. This matrix follows **Dale's** law, discovered by Henry Dale: neurons release a unique type of neurotransmitter, which means to a first approximation that the effect of one cell on its neighbours is only excitatory or only inhibitory, independently of the identity of the post-synaptic cells. In practice for our connectivity matrices, it means that columns have a constant sign. Matrices that follow Dale's law are always non-normal.

Murphy and Miller propose a very generic functional interpretation of this network: as every biologically inspired network follows Dale's rule, and because a inhibition dominated connectivity matrix that follows Dale's rule is *strongly* non-normal, the link  $t$  between the **difference** mode and the **uniform** mode is always present and strong. Indeed, for a inhibition dominated network, this link depends on the sum of the total amount (the absolute value) of the excitation and of the inhibition (see figure 3.6).

Dynamically, the Schur decomposition highlights that a difference in the level of excitation and inhibition -the amplification of the difference pattern- triggers the amplification of the uniform pattern of activity (see appendix 3.10.1).

### 3.3.1.2 Low dimensional (2D) EI networks tend to be positively correlated

Given that we are using linear dynamical systems, we can deduce easily the covariance matrix through the Lyapunov equation  $(J - I)C + C(J - I)^t = -BB^t$  (see methods 3.9.2).

As we can see in figure 3.7 B, for this example, the inhibitory population and the excitatory are positively correlated. In the appendix 3.10.1, we demonstrate mathematically that this result is very general for inhibition dominated networks.

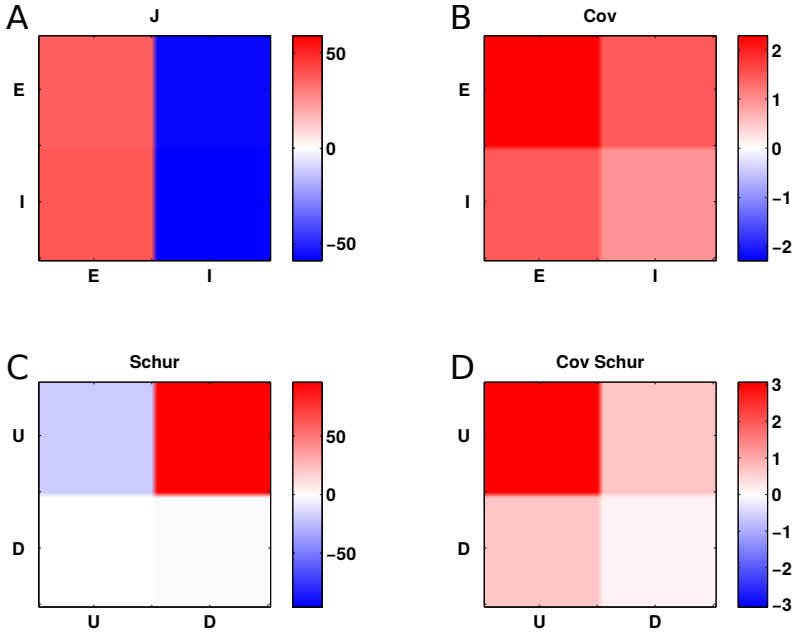


Figure 3.7: A: Connectivity matrix. B: Covariance matrix. C: Connectivity matrix in the Schur basis. D: Covariance matrix in the Schur basis. Parameters:  $J_E = 37$ ,  $J_I = 39$ ,  $\Delta = -20$  (see appendix 3.10.1).

Overall, non-normal amplification seems to be a generic mechanism capable of generating amplification. However, the problem of this E-I network is that it doesn't generate competitive amplification: we just showed that in the low dimensional rate model of E-I balanced networks, the excitatory and the inhibitory population tend to be positively correlated.

### 3.3.2 Normal competitive amplification (NCA)

#### 3.3.2.1 The symmetric competitive network in the canonical basis

Given that we found low dimensional dynamics in the data, we looked after simple low dimensional connectivity models. After the network with two populations of excitatory and inhibitory neurons, the next model in order of complexity, is a model with two populations of excitatory cells and one population of inhibitory cells, all recurrently connected. The state vector of firing rates is:  $X = \begin{pmatrix} x_{E_1}(t) \\ x_{E_2}(t) \\ x_I(t) \end{pmatrix}$ . Adopting a parametrisation similar to the one we use in the appendix 3.10.1, the connectivity matrix is:

$$J = \begin{pmatrix} \frac{J_E + \epsilon}{2} & \frac{J_E - \epsilon}{2} & -\Delta - J_E \\ \frac{J_E - \epsilon}{2} & \frac{J_E + \epsilon}{2} & -\Delta - J_E \\ \frac{J_I}{2} & \frac{J_I}{2} & -\Delta - J_I \end{pmatrix}$$

In this architecture, we split the excitatory population in the previous 2D model into two subpopulations, and we introduce an asymmetry  $\epsilon$  between the magnitude of the excitatory connections within a subpopulation and across subpopulations, in such a way that within connectivity is always a bit bigger. Note that the excitatory-excitatory part of the connectivity is symmetric, or said otherwise,  $E_1$  and  $E_2$  are fully exchangeable.

This network conserves many features with respect to the two dimensional E-I network: the total amount of excitatory post-synaptic currents received by the inhibitory population ( $J_I$ ), the total amount of inhibition received by each excitatory population ( $-\Delta - J_E$ ), and finally the total amount of excitation and inhibition received by each of the three populations ( $-\Delta$ ).

$J$  has three eigenvectors and eigenvalues:

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \lambda_1 = \epsilon \quad v_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \lambda_2 = -\Delta$$



$$v_3 = \frac{1}{\sqrt{2k^2+1}} \begin{pmatrix} k \\ k \\ 1 \end{pmatrix} \quad k = \frac{J_E + \Delta}{J_I} \quad \lambda_3 = J_E - J_I$$

$v_2$  and  $v_3$  are expanded versions of the two eigenvectors of the E-I network. We can diagonalize  $J$  using the change of basis matrix:

$$V = \begin{pmatrix} v_1, v_2, v_3 \end{pmatrix}$$

$$D = V^{-1} \cdot J \cdot V = \begin{pmatrix} \epsilon & 0 & 0 \\ 0 & -\Delta & 0 \\ 0 & 0 & J_E - J_I \end{pmatrix}$$

Because  $\epsilon$  is an eigenvalue of the connectivity matrix, by putting it close to  $1^-$ , we generate normal amplification. Several studies [160],[155], [129], [128] [166], have shown that this normal amplification is competitive: as we increase the parameter  $\epsilon$ , the two populations tend to accumulate more variance, because they are self connected in a stronger way. Also, as  $\epsilon$  increases, these two populations directly excite themselves less and less ( $\frac{J_E - \epsilon}{2}$ ), and start interacting mainly through mutual inhibition. Then, when one excitatory population is more active than its average, it tends to suppress the other population. As these populations are fed with noise, and there is no winner take all mechanisms in this network, we see alternated anti-correlated activity between  $E_1$  and  $E_2$ . Also, the uniform eigenvector  $v_2$  can be suppressed by setting the total amount of excitation plus inhibition ( $\Delta$ ) large enough, so that the competitive direction  $v_1$  can be the dominant direction by a large margin.

### 3.3.2.2 The symmetric competitive network in the Schur basis

Having a better intuition of what is non-normal amplification, we are going to revisit the NCA circuit in the Schur basis. The Gram-Schmidt procedure depends on the order that we choose the eigenvectors, but because  $v_1$  and  $v_2$  are already orthogonal, we choose these two first, and then we orthonormalize  $v_3$ , and obtain the Schur basis  $\{u_1, u_2, u_3\}$ :

$$u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad u_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad u_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

These vectors are respectively the **competitive**, the uniform and the difference. In order to rotate the connectivity matrix into the Schur basis, we use the unitary matrix  $P$ :

$$P = [u_1, u_2, u_3] = \left[ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \right]$$

And then the matrix of normal competition  $J$  can be expressed in the Schur basis as:

$$T = P^{-1}JP = P^t . J . P = \begin{pmatrix} \epsilon & 0 & 0 \\ 0 & -\Delta & \frac{(2\Delta+2J_E+J_I)}{\sqrt{2}} \\ 0 & 0 & J_E - J_I \end{pmatrix}$$

This simple transformation is very informative, because it explains the competition in a very parsimonious way. Note that, even though the connectivity matrix is strongly non-normal (as has to be the case if the connectivity respects Dale's law), the competitive mode does not receive any feed-forward links. Thus, all the amplification of competitive fluctuations in this model is purely normal. Because of this, we will refer to models of this type as producing **normal competitive amplification** (NCA). We note that this decomposition was made independently by Schaub and colleagues [132].

As Murphy and Miller predicted, in this network the non-normal amplification between the difference and the uniform mode is strong ( $\frac{2\Delta+2J_E+J_I}{\sqrt{2}}$ ). However, this non-normal amplification doesn't concerns the competitive mode.

#### 3.3.3 Transient competitive amplification (TCA)

We are interested in investigating the nature of the correlation structure when the fluctuations in the competitive mode are generated as a result of a feed-forward link from one of the other activity modes in the network as opposed to positive feed-back from itself, like in the NCA network. To that end, we studied the connectivity of networks where, in the functional (Schur) representation, we add a feed-forward link from the uniform either the difference to the competitive mode.

As we know how to build a connectivity matrix from the Schur basis to the canonical basis, we can then observe the effect of adding a feed forward link into from the uniform to the competitive mode.

Using the change of basis formula, we can proceed backwards  $J = P.T.P^t$ , and deduce how is the connectivity matrix in the canonical basis, if we know the matrix in the Schur basis. In order to keep the numbers of parameters low, we only add one feed-forward link, and we choose it to be the link from the uniform to the competitive.

$$T = \begin{pmatrix} \lambda_C & M & 0 \\ 0 & \lambda_U & V \\ 0 & 0 & \lambda_D \end{pmatrix} \quad P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} \end{pmatrix} \quad J = P.T.P'$$

$$J = \frac{1}{2} \begin{pmatrix} \frac{\lambda_D + \sqrt{2}.V + \sqrt{6}.M + 2.\lambda_U + 3.\lambda_C}{3} & \frac{\lambda_D + \sqrt{2}.V + \sqrt{6}.M + 2.\lambda_U - 3.\lambda_C}{3} & \frac{-2.\lambda_D - 2.\sqrt{2}.V + \sqrt{6}.M + 2.\lambda_U}{3} \\ \frac{\lambda_D + \sqrt{2}.V - \sqrt{6}.M + 2.\lambda_U - 3.\lambda_C}{3} & \frac{\lambda_D + \sqrt{2}.V - \sqrt{6}.M + 2.\lambda_U + 3.\lambda_C}{3} & \frac{-2.\lambda_D - 2.\sqrt{2}.V - \sqrt{6}.M + 2.\lambda_U}{3} \\ \frac{-2.\lambda_D + \sqrt{2}.V + 2.\lambda_U}{3} & \frac{-2.\lambda_D + \sqrt{2}.V + 2.\lambda_U}{3} & \frac{4.\lambda_D - 2.\sqrt{2}.V + 2.\lambda_U}{3} \end{pmatrix}$$

As we did in the appendix 3.10.1, we can re-parametrize the matrix in the canonical basis :

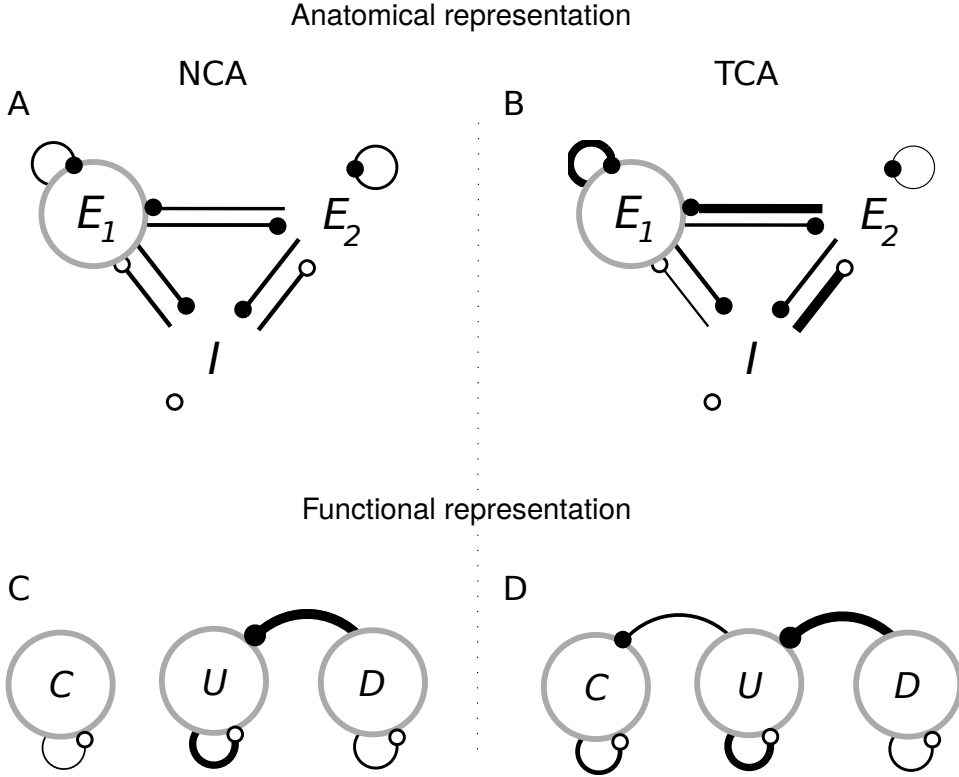
$$J = \begin{pmatrix} \frac{J_E + \epsilon + \psi}{2} & \frac{J_E - \epsilon + \psi}{2} & -\Delta - J_E + \frac{\psi}{2} \\ \frac{J_E - \epsilon - \psi}{2} & \frac{J_E + \epsilon - \psi}{2} & -\Delta - J_E - \frac{\psi}{2} \\ \frac{J_I}{2} & \frac{J_I}{2} & -\Delta - J_I \end{pmatrix}$$

$$J_E = \frac{\lambda_D + \sqrt{2}.V + 2.\lambda_U}{3}, \quad J_I = \frac{-2\lambda_D + \sqrt{2}.V + 2.\lambda_U}{3}, \quad J_E - J_I = \lambda_D$$

$$\epsilon = \lambda_C, \quad \psi = \frac{M\sqrt{6}}{3}, \quad \Delta = -\lambda_U$$

We see that we obtain a matrix, whose particularity with respect to the normal competitive one is that it has an asymmetry in the connectivity of the two excitatory populations.

Observing the connectivity matrix of both models in the anatomical basis may lead to assume that there is just a quantitative difference in the way these circuits operate, whereas in the functional representation it clearly appears that both circuits behave in a qualitatively different manner. The competitive pattern is amplified in two different ways: in NCA, the negative feedback is small  $|\lambda_c - 1| \ll 1$ , whereas in TCA, the negative feedback



*Figure 3.8: Anatomical and functional representation of normal competitive amplification (A,C) and transient competitive amplification (B,D).*

of the competitive mode is bigger, but this mode also receives feed-forward input from the uniform mode.

Given the way we built this matrix, this model allows to have an alternative way of generating competitive amplification: purely through normal amplification or through non-normal amplification. Using non-normal amplification, we can trade pure normal amplification against non-normal amplification. It might appear like NCA is a particular case of TCA, and it is indeed the case. However, because it uses a new mechanism of generating amplification, that as we will see generates different experimental predictions, we consider them as two different models.

### 3.4 Model predictions and comparison with the data

In general, the problem of deducing the dynamics from the connectivity is a hard problem - which is not solved in spiking neurons for example. In the case of linear dynamical systems, this problem is more tractable, and we have two tools at our disposal: first the Schur decomposition, which as we said allows to have a good qualitative idea about the system dynamics. Second, given a linear system determined by the equation:  $\dot{X} = (J - I)X + B\xi$ , we can obtain the correlation matrix  $C$ , solving the Lyapunov equation:

$$(J - I)C + C(J - I)^t = -BB^t$$

For very large systems, the analytical expression of the correlations given by the Lyapunov equation is intractable, because there are too many terms. However, in low dimension, we can use both the Schur decomposition and the analytical solutions to understand better our networks. For instance, in the NCA (figure 3.8 C), because the competitive mode is isolated we can deduce directly that the covariance between the competitive and the uniform mode is nil. Also, when we have a link (of positive strength) between two modes, we can directly state that they are positively correlated like the difference and the uniform.

In the subsequent paragraphs, we will use these two tools to infer properties of the correlations for the NCA and for the TCA.

#### 3.4.1 NCA and TCA generate negative correlations among $E_1$ and $E_2$

Using the change of basis formula from the Schur basis to the canonical basis (see appendix 3.10.3), we are first going to prove that the competitive mode has a decorrelating effect on the correlation between  $E_1$  and  $E_2$ . We remind that we call  $u$  the uniform mode  $(1, 1, 1)$ ,  $c$  the competitive mode  $(1, -1, 0)$  and  $d$  the difference mode  $(1, 1, -2)$ .

The covariance between  $E_1$  and  $E_2$  is given by the term:

$$C_{(E_1, E_2)} = \frac{c_{uu}}{3} - \frac{c_{cc}}{2} + \frac{c_{dd}}{6} + \frac{\sqrt{2}c_{ud}}{3}$$

Considering the Schur architecture, it is clear that  $c_{UD}$  is a strictly positive quantity. By definition, the autocovariances, or variances,  $c_{ii}$  are also positive. We can then conclude that the variance of the competitive mode has a decorrelating effect on the covariance between  $E_1$  and  $E_2$ .

### 3.4.2 Asymmetry in the connectivity generate difference in the variances of $E_1$ and of $E_2$

We can also reuse the same arguments as before, and consider the variances of  $E_1$  and  $E_2$ :

$$\begin{aligned} \text{var}(E_1) = C_{E_1 E_1} &= \frac{c_{cc}}{2} + \frac{c_{uu}}{3} + \frac{c_{dd}}{6} + \frac{\sqrt{6}c_{cu}}{3} + \frac{\sqrt{3}c_{cd}}{3} + \frac{\sqrt{2}c_{ud}}{3} \\ \text{var}(E_2) = C_{E_2 E_2} &= \frac{c_{cc}}{2} + \frac{c_{uu}}{3} + \frac{c_{dd}}{6} - \frac{\sqrt{6}c_{cu}}{3} - \frac{\sqrt{3}c_{cd}}{3} + \frac{\sqrt{2}c_{ud}}{3} \end{aligned}$$

In the NCA model, the Schur basis, the competitive mode is isolated from the uniform and from the difference mode, which makes that their covariance  $c_{cu}$  and  $c_{cd}$  be nil and therefore it is expected from these equations that both populations have the same variance, as we could guess from the symmetry of the connectivity.

In TCA, the asymmetry in the connectivity induces a link between the competitive mode and the uniform that correlates positively the competitive, the uniform and the difference mode:  $c_{cu} > 0$  and  $c_{cd} > 0$ . The variance of  $E_1$  in the TCA model is equal to the variance in NCA to which we add a positive constant ( $\frac{\sqrt{6}c_{cu}}{3} + \frac{\sqrt{3}c_{cd}}{3}$ ), whereas the variance of  $E_2$  is equal to the variance of NCA minus that constant.

### 3.4.3 Asymmetry in the connectivity generate difference in covariances between the two populations and the inhibitory neurons

The covariance between  $E_1$  and I is:

$$C_{(E_1, I)} = \frac{c_{ui}}{3} - \frac{c_{di}}{3} + \frac{\sqrt{6}c_{cu}}{6} - \frac{\sqrt{3}c_{cd}}{3} - \frac{\sqrt{2}c_{ud}}{6}$$

and the covariance between  $E_2$  and I is:

$$C_{(E_2, I)} = \frac{c_{uu}}{3} - \frac{c_{dd}}{3} - \frac{\sqrt{6}c_{cu}}{6} + \frac{\sqrt{3}c_{cd}}{3} - \frac{\sqrt{2}c_{ud}}{6}$$

Once again, the feed-forward link in the Schur basis correlates the competitive with the uniform and the competitive with the difference mode, and this creates an asymmetry in the way both populations correlate positively with I. However, the effects cancel out at least partially in each covariance:  $\frac{\sqrt{6}c_{cu}}{6} - \frac{\sqrt{3}c_{cd}}{3}$  and  $-\frac{\sqrt{6}c_{cu}}{6} + \frac{\sqrt{3}c_{cd}}{3}$ . Even if we add systematically a constant to the covariation of  $cov(E_1, I)$  with respect to the covariation of the NCA case, and we subtract this constant to the covariation of  $cov(E_2, I)$ , it might be that this difference in covariance is less salient than the other two (see figure 3.9 C, D).

#### 3.4.4 NCA crosscorrelogram is symmetric while TCA has a lag in the crosscorrelogram

The crosscorrelogram of NCA between  $E_1$  and  $E_2$  is symmetric around 0: the delayed correlation between them is negative at zero lag and tends to zero as the lag increases, and this in a similar manner for both populations:  $\langle x_{E_1}(t).x_{E_2}(t + \tau) \rangle = \langle x_{E_2}(t).x_{E_1}(t + \tau) \rangle$  (figure 3.9 G).

In the case of TCA (see figure 3.9 H), there is a lag in the cross-correlogram between  $E_1$  and  $E_2$ , in such a way that an excess of firing of  $E_1$  -the most self-connected population- precedes an excess of silence of  $E_2$  (see figure 3.9 G,H).

The non-normal mechanism which is at the origin of non-contractive dynamics in the mean trajectories is also responsible for this non-monotonical crosscorrelogram (see methods 3.9.3).

Predictably, due to the fact that NCA proceeds by integrating the noise, the time scale of the competition is much slower in NCA than in TCA.

#### 3.4.5 Angle between the PC1 and the uniform

The  $PC_1$  in the case of NCA is orthogonal to the uniform, whereas we see that in the case of the TCA it is less orthogonal (see figure 3.9 E,F).

In the NCA, all the parametrisations of the connectivity matrix lead to a Schur decomposition with the three orthogonal modes: competitive,

uniform and difference. However, not all the parametrisations of the NCA connectivity matrix lead to competition (see paragraph normal competitive amplification). NCA's, Schur decomposition has one isolated competitive mode and a feed-forward difference to uniform mode. To generate competition, the variance associated with the competitive mode must overcome by far the amplification generated by the uniform mode. We achieve this by making the negative feed-back of the competitive mode very small. As  $\epsilon \rightarrow 1^-$ , the  $PC_1$  aligns perfectly with the competitive mode.

In the TCA case however, the negative feedback of the competitive mode is bigger, so the part of the variance of the competitive mode that is due to the negative feed-back is smaller. A bigger part of the variance is inherited from the uniform mode and then even if the activity is dominated by the competitive mode, it has a much bigger projection onto the uniform mode. Therefore, for TCA, for a similar level of amplification, the  $PC_1$  is less orthogonal to the uniform mode in comparison with the  $PC_1$  of the NCA.

The choice of parameters was only based on the two following constraints: a) finding a set of parameters for the NCA and for the TCA which produces a similar amount of variance. b) make sure that the connectivity matrix is Dale.

#### 3.4.6 Comparison with data

In the figure 3.10, we plotted the mean cross-correlograms for every experiment (see methods 3.9.4). For every experiment, we label as  $E_1$  (resp.  $E_2$ ) the group which has more variance. The cross-correlogram between  $E_1$  and  $E_2$  is plotted as it is.

As we see, in nearly every experiment there is a systematic lag in the ccg of  $E_1$  and  $E_2$ : excess of firing of  $E_1$  precedes an excess of silence of  $E_2$ . This analysis suggest that this prediction is a signature of non-normal amplification.



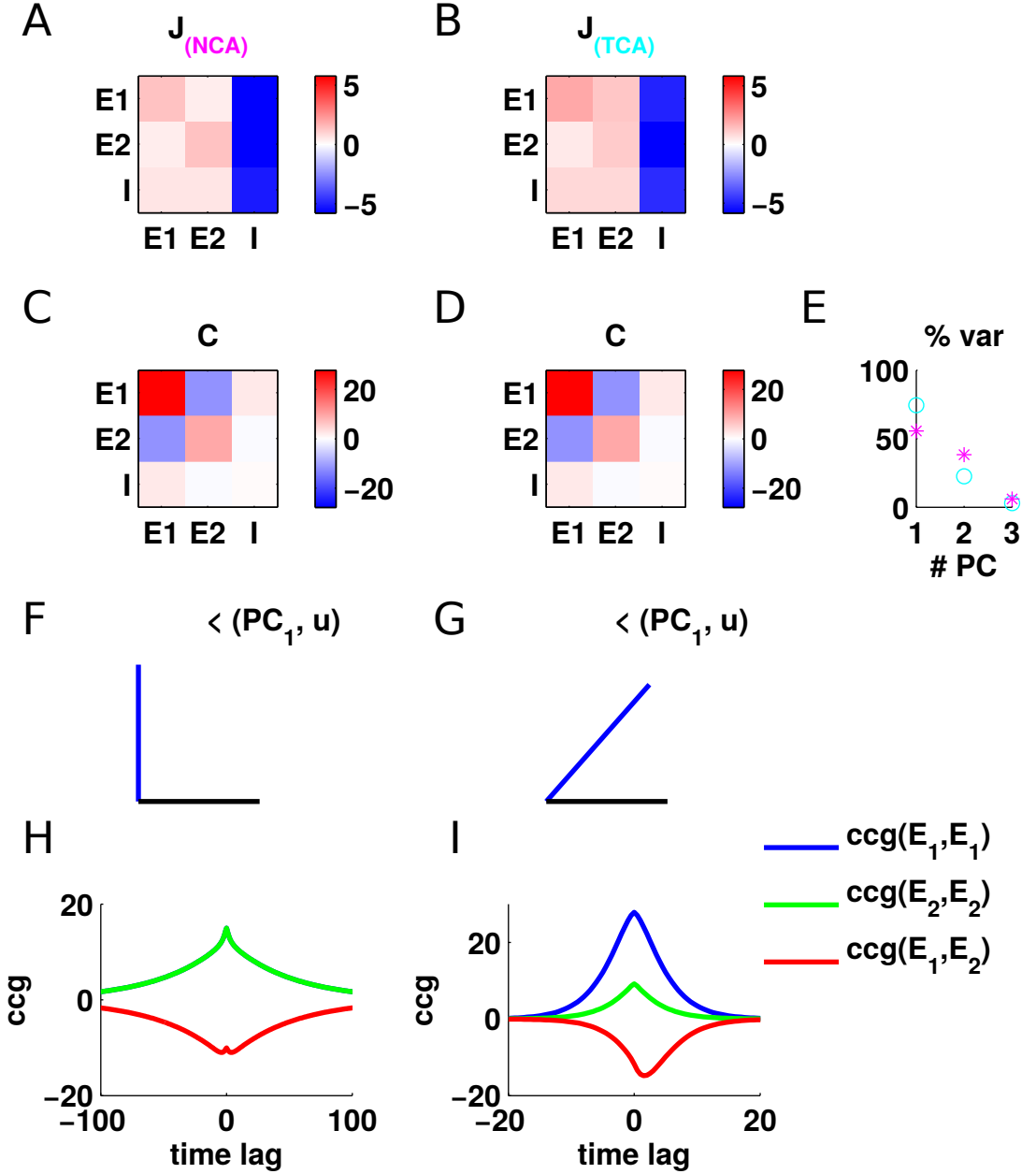


Figure 3.9: Predictions from NCA and TCA: A, B: Connectivity matrices. C, D: covariance matrix. E: % of variance. F, G: angle of the first PC with the uniform. H, I: ccgs. Parameters NCA: ( $\lambda_c = 0.98, \lambda_u = -3, \lambda_d = 0.6, M = 0, V = 9, \tau = 1$ ). Parameters TCA: ( $\lambda_c = 0.69, \lambda_u = -3, \lambda_d = 0.6, M = 1, V = 9, \tau = 1$ ).

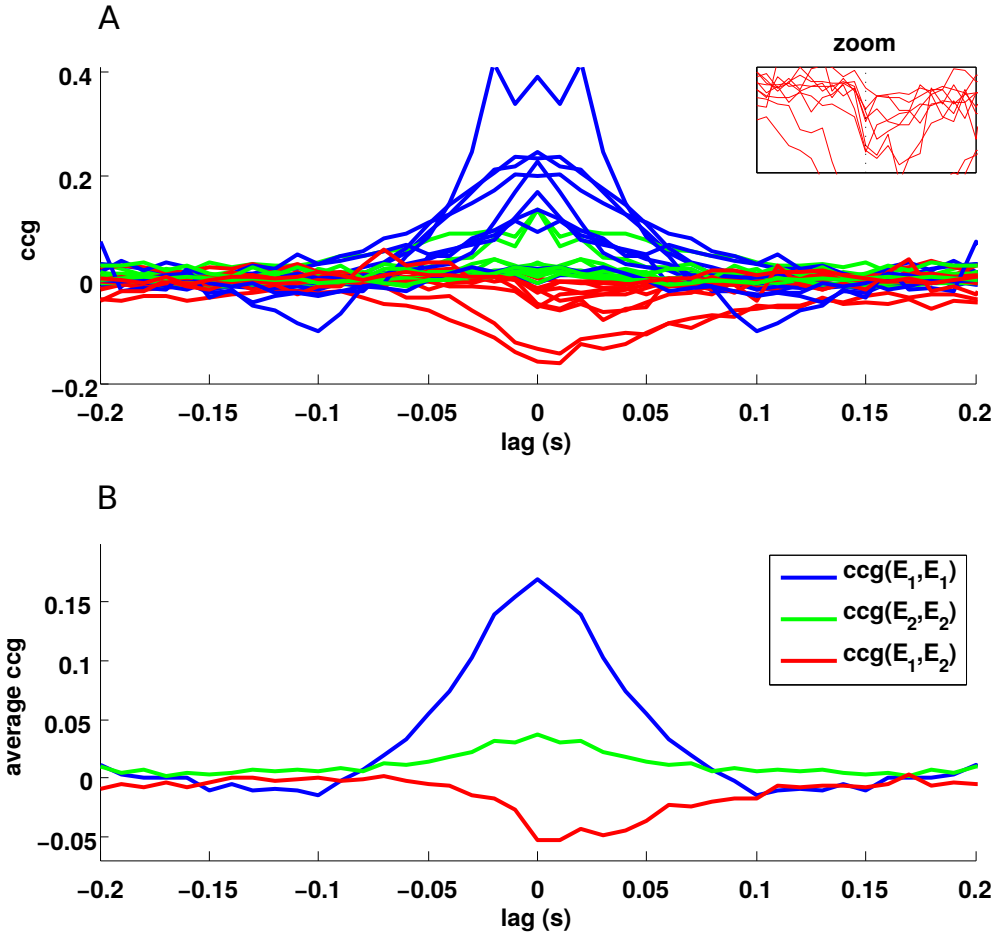


Figure 3.10: (A): Average cross-correlograms computed for every experiment. Each experiment constitutes a triplet  $ccg(E_1, E_1), ccg(E_2, E_2), ccg(E_1, E_2)$ . On the lower panel (B), we average all the ccgs accross experiments.

## 3.5 Generating competitive low dimensional dynamics in a high dimensional network

### 3.5.1 Limitations of the low dimensional motif

The question we are going to answer in this section is how can a neural network, which in principle could have huge degrees of freedom in the connectivity, generate low dimensional dynamics.

In the second chapter we discovered that the apparent unstructured activity during the active state is competitive. One of the most salient features of the competitive activity is that it is low dimensional. We modelled the firing rates and the covariances using a three dimensional linear dynamical system, and reproduced some of the observables, notably the difference in the variance across the populations, the direction of the PC1 with respect to the uniform and also the lag in the crosscorrelogram between  $E_1$  and  $E_2$ .

Nevertheless, the proposed model has limitations: what we call the population rate in our 3D model is the mean firing rate of the cells in a population. Still, in the data we have abundant heterogeneity and the distinction between the two excitatory populations is not discrete. Although approximately half of the neurons has positive loadings and the other half has negative loadings, the neurons with small loadings in absolute value are almost uncorrelated between them, whereas pairs of neurons with high loadings in absolute value have a strong correlation. In the data, many of the observables are spread: the correlation matrix, the histogram of pairwise correlation, the eigen-spectrum, the PC loadings, and so on. One of the advantages to go to high dimensions is that, for strongly connected networks with strong quenching of the uniform activity, the mean correlations go as  $1/N$ , where  $N$  is the size of the network [123], [64].

In what follows, we will explain how can a high dimensional circuit generate competitive low dimensional dynamics and how to introduce a gradation in the competitive activity.

### 3.5.2 Generating an expanded high dimensional low-rank matrix

One possibility to generate low dimensional dynamics is that the connectivity matrix has low **rank**, i.e. few degrees of freedom. For instance, consider a matrix equal to  $\vec{v}.\vec{v}^t$ : the outer product of two vectors has rank one and has only one eigenvector:  $\vec{v}$ .

Expanding the low dimensional TCA model (3.11 A) in high dimension is going to result in a square matrix of size N and of rank equal to 3 (figure 3.11 B).

The appropriate way of stretching the connectivity matrix from low dimension to high dimension, is going from the Schur basis in low dimension to the canonical basis in high dimension, with high dimensional change of basis vectors. In three dimensions, the Schur vectors are:

$$u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad u_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad u_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

And in N dimensions, the Schur vectors are:

$$u_1 = \frac{1}{\sqrt{f.N}} \begin{pmatrix} 1 \\ \dots \\ 1 \\ -1 \\ \dots \\ -1 \\ 0 \\ \dots \\ 0 \end{pmatrix} \quad u_2 = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 \\ 1 \\ \dots \\ \dots \\ \dots \\ 1 \\ 1 \end{pmatrix} \quad u_3 = \frac{1}{\sqrt{f + (1-f)L^2}} \begin{pmatrix} 1 \\ \dots \\ 1 \\ 1 \\ \dots \\ 1 \\ -L \\ \dots \\ -L \end{pmatrix}$$

$L = \frac{f}{1-f}$ . The proportion of excitatory neurons is  $f$ . The difference vector is modified such that  $\sum_i u_{3|i} = 0$ .

The stretched connectivity matrix in high dimension is:

$$M = \sqrt{N} \cdot [u_1, u_2, u_3] \cdot T \cdot [u_1, u_2, u_3]^t$$

T being still the 3 by 3 Schur decomposition of the low dimensional model. The dimensions of M are then  $\langle N.3 \rangle \cdot \langle 3.3 \rangle \cdot \langle 3.N \rangle = \langle N.N \rangle$ .

Supposing that the elements in  $T$  are  $O(1)$ , in order to be in the balanced state [123], we choose a scaling of  $\sqrt{N}.T$ : the connections in the canonical basis scale as  $O(\frac{1}{\sqrt{N}})$  and the eigenvalues as  $O(\sqrt{N})$ .

It is patent that building the connectivity matrix this way leads to a high dimensional system which has the same defects as the low dimensional system, because groups remain homogeneous so that there is not going to be gradation in the covariances of neurons (see figure 3.11 B).

The solution of this issue is to introduce a graded vector that follows a linear progression (e.g: 1, 0.9, 0.8, ..., 0.1, 0) from a one towards zero for the first group of excitatory neurons and then from zero to -1 for the second group of excitatory neurons. To allow many having different progressions towards zero, we elevate this linear progression to a certain power  $\kappa$ . We then re-normalize the vector.

$$u_1 = \frac{1}{||u_1||} \begin{pmatrix} 1^\kappa \\ 0.9^\kappa \\ \dots \\ 0 \\ 0 \\ \dots \\ -0.9^\kappa \\ -1^\kappa \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

This solution leads to a low rank matrix with graded connections between the neurons. As shown by the spectral decomposition (see appendix 3.10.4), the contribution to the connectivity of the competitive vector is given by the sum of the two matrices:  $\lambda_c.u_1u_1' + t_{12}u_1.u_2'$ .

In figure 3.11 we can see an example of three connectivity matrices: the low dimensional model of three populations (A), and two matrices of 2383 neurons without graded competitive vector (B) and with graded competitive vector (C). In (D), we added noise in the connectivity, in a principled way that we will explain in a subsequent paragraph.

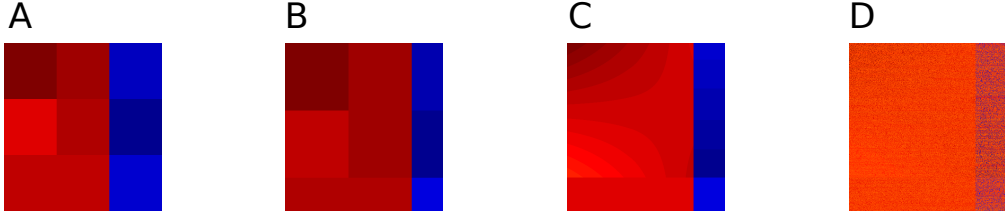


Figure 3.11: Example of three connectivity matrices: A: low dimensional matrix, B: high dimensional non graded matrix and C: high dimensional graded matrix. D: high dimensional graded matrix with noise on the connectivity. The inhibitory connections are depicted in blue and the excitatory connections in red and orange. The color maps are not the same colorbars in order to try to optimise the discrimination of the details. The high dimensional matrices are Dale. Parameters: ( $\lambda_c = 0.69, \lambda_u = -3, \lambda_d = 0.6, M = 1, V = 19, f = 0.8, \kappa = 1.05, R = 0.79, \kappa = 1.05, N = 2383, p = 0.2$ ).

### 3.5.2.1 Correspondence between the covariance of the low dimensional model and of the high dimensional network (without noise on the connectivity)

In the appendix 3.10.3, we show that the covariance also follows a change of base rule. When we expand our low dimensional system into a higher dimensional system, this change of basis also holds. The  $N$  by  $N$  covariance matrix  $C$  -in the canonical basis- is the expansion into high dimension of the 3 by 3 covariance matrix -in the Schur basis  $C_s$ .

$$C = \frac{1}{\sqrt{N}} [u_1, u_2, u_3] \cdot C_s \cdot [u_1, u_2, u_3]^t$$

### 3.5.3 Adding noise to the connectivity

The matrix  $M$  is a high dimensional version of the TCA model. However, by construction, the connectivity matrix is low-rank, and the eigenvectors only explore a small 3 dimensional subspace in an  $N$  dimensional space. In this section we see to add noise to the connectivity matrix in a principled way that preserves the original 3 eigenvalues and eigenvectors, but that adds  $N-3$  others. This allows to have a full rank connectivity matrix which in principle could amplify the activity in different directions. The noise also

adds heterogeneity in the connectivity between the neurons (figure 3.11 D).

We propose a high dimensional model that includes two terms: a low rank deterministic matrix  $M$  (which is a stretched version of the low dimensional model), plus a noise matrix  $F$ .

Using a result from Rajan and Abbott [122], we established a condition that ensures that we can superpose the Schur decomposition of the sum matrix  $J = M + F$  as the sum of the Schur decompositions of  $M$  and  $F$  (see appendix 3.10.7). This decomposition gives a dynamical interpretation of the noisy connectivity patterns, and allows to understand how the random connectivity patterns interact with the deterministic patterns.

As we see in figure 3.12, the connectivity matrix can be rotated to the Schur basis. As Hennequin [64], we understand that the noise in the connectivity also adds directions of amplification in the Schur basis, and that given how we rotated our connectivity basis from the canonical basis to the Schur basis, the variance from the noisy modes flows to the deterministic modes in a feed-forward manner and at the same time, it is quenched in every node.

The condition that we impose on the noisy matrix is valid for any kind of noise we want to use to create heterogeneity in the connections. As long as the noisy connections are drawn independently from a distribution, the eigenspectrum of the noisy connectivity matrix will be situated in a circle centered in zero on the complex plane, this is called Girko's circle law (see [122],[2]).

In the appendix 3.10.7, we study two types of noise in the connectivity: Gaussian noise and Bernoulli noise. Gaussian noise consists simply in sampling the elements of the matrix  $F$  from a Gaussian distribution of zero mean and standard deviation  $\frac{R}{\sqrt{N}}$ :  $F_{i,j} \sim N(0, \frac{R}{\sqrt{N}})$ , where  $N$  is the size of the network and  $R$  is the radius of the circle of eigenvalues in the complex plane.

The advantage of using Gaussian noise is that the spectral radius  $R$  can be fixed independently of the network size and therefore, we can build networks of the size we want. If the spectral radius is above 1 -the decay rate in the differential equation  $\dot{X} = -X + JX + B\xi$ -, then the system becomes unstable and explodes, which only means in practice that our modelling assumptions are not correct anymore.

The problem of the Gaussian noise is that: first, it can lead to a break of Dale's rules -to which we compel ourselves to work in-. Second is that it leads to have a fully connected recurrent network, which is not that biologically

plausible, because we know that in cortex neurons have a probability of being connected that is below 50%.

We then turned to use Bernoulli noise. Up to some technical subtleties, we build the noise matrix by sampling connections from a copy of the deterministic matrix: for each connection, we sample a random number between 0 and 1. If the random number is smaller than a certain finite probability  $p = 0.2$ , we set a connection, and this connection is of the same order of magnitude as the connection on the deterministic matrix. If the random number is higher than the probability of connection, there is no connection established. The problem of using Bernoulli noise is that the mean  $p$  and the variance  $p(1 - p)$  of the "coin" are linked. Therefore, the spectral radius increases with the network size. In the appendix [3.10.7](#), we explain how to re-normalize the connectivity matrix in such a way that the spectral radius is equal to a certain value  $R < 1$ . The consequence of this normalization is however that we can not choose the network size of our network. Once we fix the probability of connection, the initial strength of the connections and the probability of connections, we obtain a finite network size.

The results that we are going to examine at present, are done with Bernoulli noise.



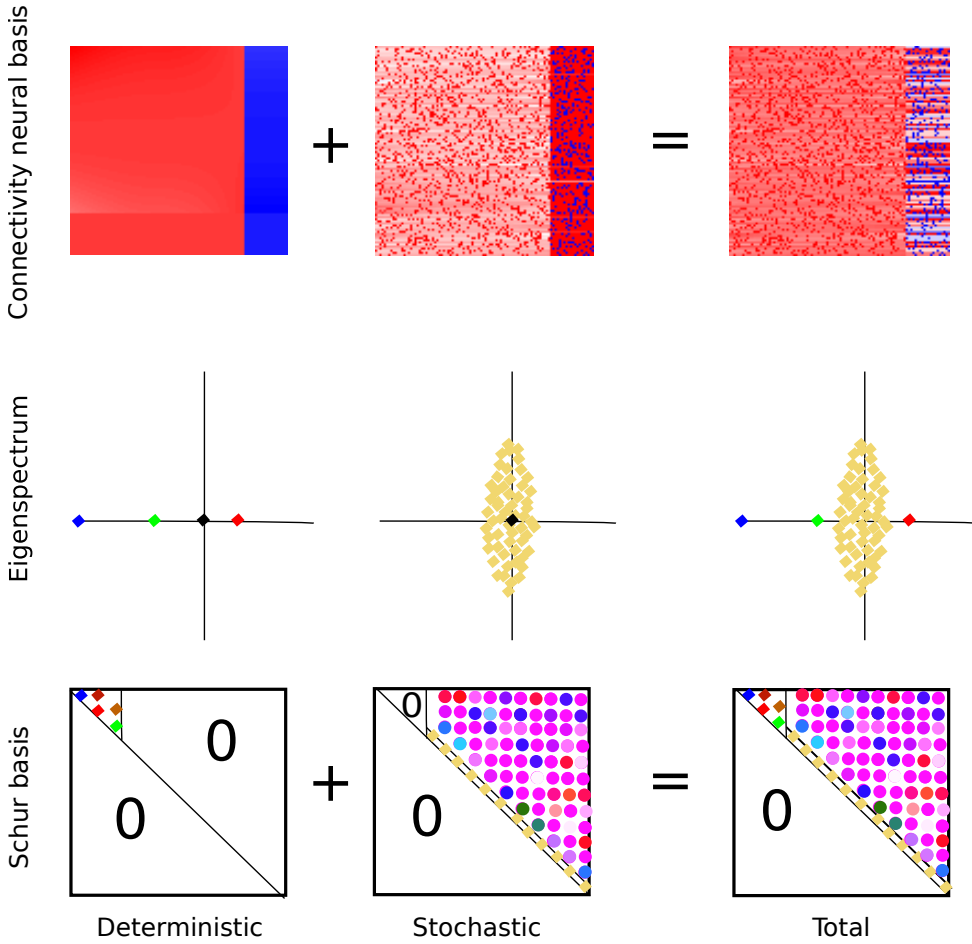


Figure 3.12: Top panel: decomposition of the total connectivity matrix, as sum of the deterministic part and the stochastic part. In the middle panel we see a cartoon of the eigenspectrum in the complex domain (ordinate real part, abscissa imaginary part) of the deterministic matrix (with 3 out of  $N$  non nil values), of the stochastic (with  $N-3$  non nil values) and finally of the eigenspectrum of the summed matrix. At the bottom we see the Schur decomposition of each of the matrices.

## 3.6 Predictions of the high dimensional model

### 3.6.1 Correspondance between the HD and the LD version of the model

In the following figure 3.13, we see the averaged cross-correlograms for the high dimensional model and for the low dimensional model.

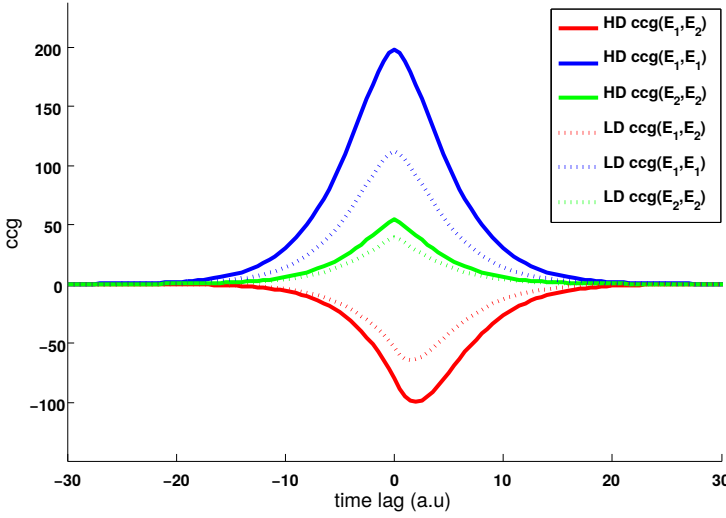


Figure 3.13: Comparing low dimensional model with high dimensional model. In dashed we see the cross-correlograms of the low dimensional model. Parameters used: ( $\lambda_c = 0.69, \lambda_u = -3, \lambda_d = 0.6, M = 1, V = 19, f = 0.8, \kappa = 1.05, R = 0.79, \kappa = 1.05, N = 2383, p = 0.2$ )

We remark that these cross-correlograms are very similar, but they don't overlap. The noisy modes contribute to the variance of the deterministic modes, so that the amplitudes of the ccgs are bigger. When we decrease the target spectral radius  $R$  towards zero, the plots overlap.

### 3.6.2 Comparison between data and model

In the following plots we will compare many observables in the high dimensional model and in the data.

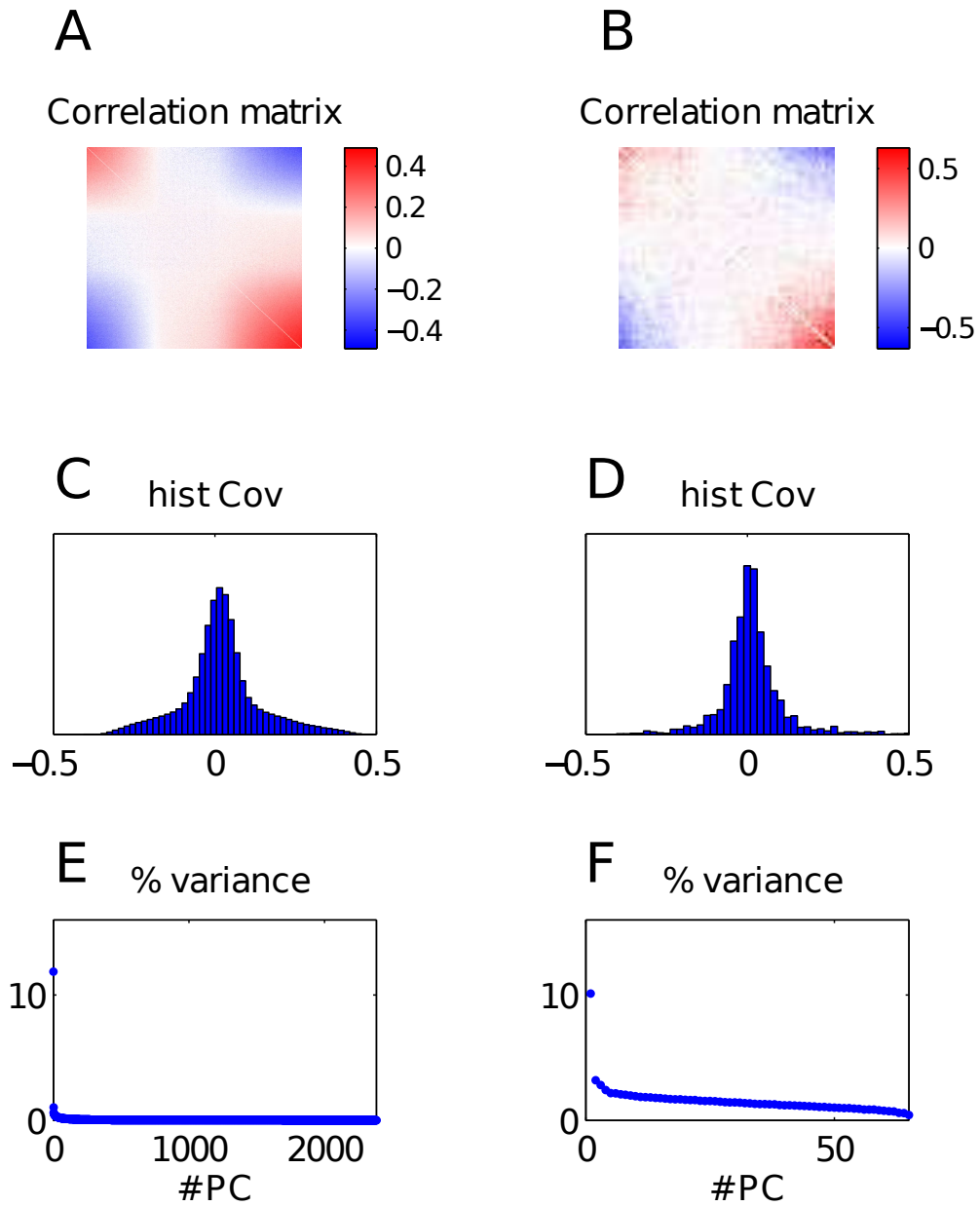


Figure 3.14: Comparison between model and data. Model: (A,C,E). Data: (B,D,F). A, B: Correlation matrices. C, D: histogram of pairwise correlations. E, F: eigenspectrum. Parameters: ( $\lambda_c = 0.69, \lambda_u = -3, \lambda_d = 0.6, M = 1, V = 19, f = 0.8, \kappa = 1.05, R = 0.79, \kappa = 1.05, N = 2383, p = 0.2$ ). The spikes trains are convolved with a mexican hat of parameters  $\sigma_1 = 0.1s, \sigma_2 = 0.4s$ .

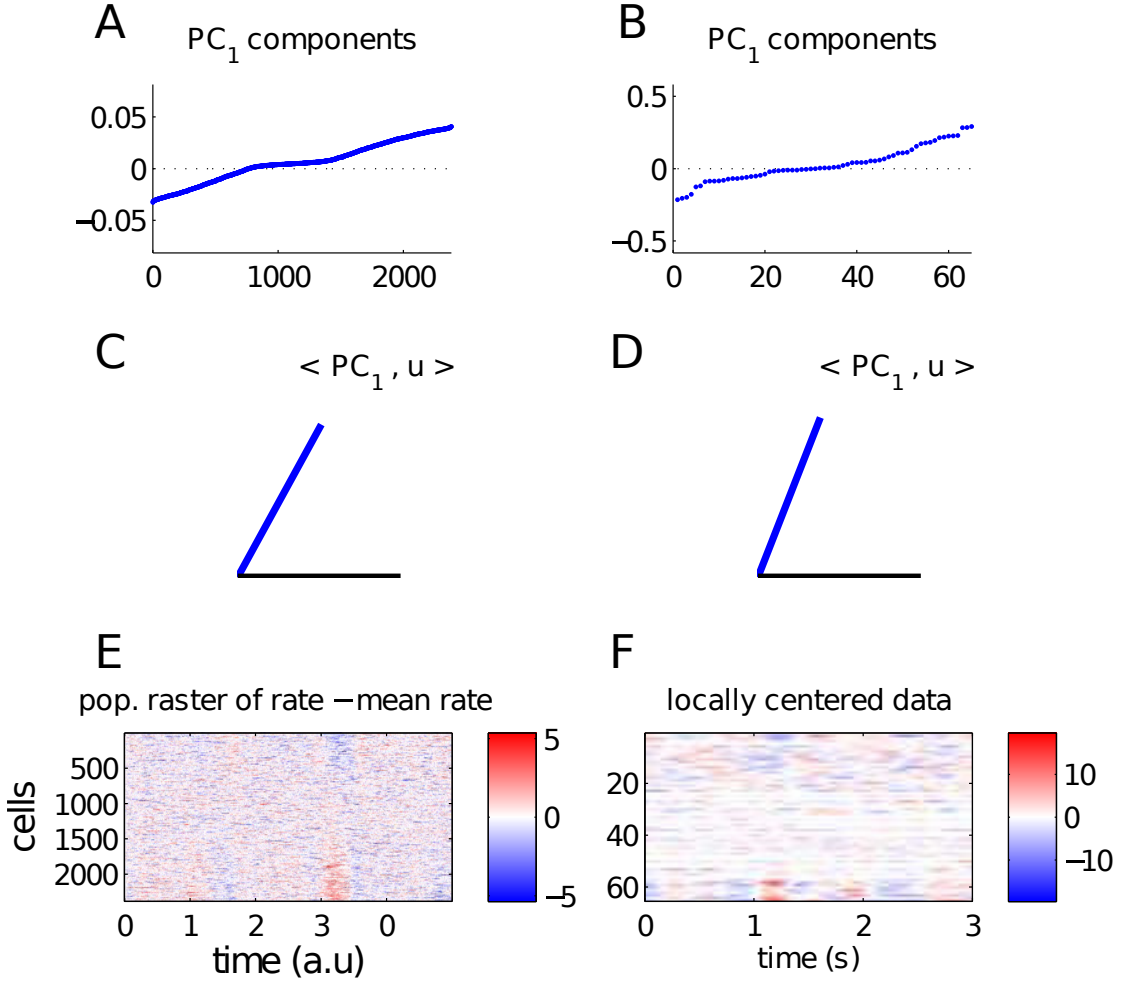


Figure 3.15: Comparison between model and data. Model: (A,C,E). Data: (B,D,F). A, B: sorted PC1 loadings. C, D: angle of the first PC with the uniform. E, F: raster plots. Parameters: ( $\lambda_c = 0.69, \lambda_u = -3, \lambda_d = 0.6, M = 1, V = 19, f = 0.8, \kappa = 1.05, R = 0.79, \kappa = 1.05, N = 2383, p = 0.2$ ). The spikes trains are convolved with a mexican hat of parameters  $\sigma_1 = 0.1s, \sigma_2 = 0.4s$ .

As we see in figure 3.14 and 3.15, there is good agreement between the model and the data.

## 3.7 Other models of competition

Along this study we have been concerned with figuring-out alternative circuit motifs that generate competition, as a way of understanding if our model is trivial or not. Even if we have tried many possibilities, we haven't come with alternative ways of generating competitive amplification, as we observe it in the data. In what follows, we will present two different circuit motifs that might generate competition, but that we discarded for reasons that we will precise.

### 3.7.1 Alternative low dimensional motif to TCA

The first TCA circuit that we studied and that we considered an ideal candidate for describing the data was not the circuit we presented in this chapter, but rather the model that we will describe here. In this circuit, instead of having a serial feed-forward chain from D to U and from U to C, we had a link from D to U and also a link from D to C:

$$T = \begin{pmatrix} \lambda_U & 0 & L \\ 0 & \lambda_C & K \\ 0 & 0 & \lambda_D \end{pmatrix}$$

We call this model UCD, because of the order of the modes. This low dimensional model is able to generate competitive amplification and makes very similar predictions as our actual TCA model.

However, when we started studying the high dimensional version of the network, we realized that the UCD model had certain characteristics that made it a worst candidate to describe competition.

The main difference between the TCA motif and the UCD, is that in the TCA model, the variance flows through all the modes to end in the competitive mode: first from the difference mode to the uniform mode and then to the competitive mode (figure 3.8). Then, in order to amplify the competitive mode, the link from the uniform to the competitive mode in the TCA model needs to be much smaller than the link from the difference to the competitive mode in the UCD model. The fact of needing a big link from the difference to the competitive makes that in the UCD model we also need a bigger link from the difference to the uniform in order to keep the matrix Dale (see appendix 3.10.4.1). Also the eigenvalue corresponding to

the uniform mode needs to be bigger in absolute value in the UCD in order to be in the dynamical regime we want. Overall, many parameters in the UCD circuit end up being bigger than in the TCA motif -up to five times for some of them-. This has two two consequences:

- The first one is a practical consequence, linked with the choice of the Bernoulli noise in the connectivity. If the coefficients in the Schur basis are too big, when we build the corresponding high dimensional network, we will have a very big spectral radius, so after re-normalisation and resizing of the network, we will end up with a network size that might be too big to solve numerically in a reasonable time. For instance, in what respects solving numerically the Lyapunov equation, there is a big difference in terms of memory resources and computation time if we are dealing with a network of 2000 neurons instead of a network of 10000 neurons, because the time of computation is quadratic in terms of the number of elements. If we had chosen a less realistic noise on the connectivity, like Gaussian noise, we would be free to choose the size of the network, but we would still have the same problem when simulating very big networks.

- The second consequence is worst, but nonetheless it is not an insurmountable obstacle. When we build a high dimensional model, introducing a graded competitive vector allows to reproduce a more realistic covariance matrix. One of the practical consequences of introducing a graded Schur eigenvector, is that the faster the progression from 1 to 0 (in the unnormalized competitive vector), the bigger maximal value of this progression will be in the normalized version of the vector. For example, lets suppose that the direction of the competitive vector is given by

$$\vec{u}_c = \left( 1^\kappa \quad 0.98^\kappa \quad \dots 0 \quad 0 \quad \dots \quad -0.98^\kappa \quad -1^\kappa \quad 0 \quad \dots 0 \right)^t$$

The bigger  $\kappa$  is, the bigger the first component of the competitive vector will be, once we normalize it  $\vec{u}_c \leftarrow \frac{\vec{u}_c}{\|\vec{u}_c\|}$ . This might break the Dale condition -to which we compel ourselves to work in (see spectral decomposition in the appendix 3.10.4.1). In practice, small deviations from Dale's rule have no consequence on the dynamics, but big breaks in Dale's rule do often generates a switch in the dynamics.

The only way to overcome the fact that the connectivity matrix is not Dale in high dimensions is simply to increase the  $D \rightarrow U$  link. This link is the one that guarantees the Dale property of the connectivity matrix. However, increasing the strength of the link D to U might makes us run

in the previously evoked problem. Finally, another disadvantage of having a too strong difference to uniform link, both in the UCD and in the TCA models, is that in the simulations we observe a discontinuity in the derivative at zero lag for the cross-correlogram of  $E_1$  and  $E_2$ , which is not apparent in the experimental data.

#### 3.7.2 Translation-invariant connectivity matrices might generate competitive activity, but with different characteristics

Because the Schur basis is orthogonal, any network where the uniform vector is part of the Schur basis is bound to have other dimensions whose dynamics, if amplified by the connectivity, would qualify as "competitive amplification". This is bound to almost always be the case in high dimensions. Of course, an exhaustive search of all the possible alternative models is impractical, but there are a certain "canonical" models which it might be interesting to investigate, in an effort to understand the extent to which they provide adequate explanations of the experimental data. In particular, we would like to know not only if they can produce competitive amplification (which as just noted is almost trivially bound to be true), but whether, as we observe in the data, competitive amplification is close to one-dimensional and dominant (fluctuations along a single competitive pattern appear to be significantly amplified). One such canonical model is a network with translational invariance, such as typically used to explain orientation or direction selectivity in the visual cortex [11]. In fact in their paper introducing non-normal amplification, Murphy and Miller [105] considered non-normality in such a network. They considered connectivity matrices such that:

$$W = \begin{pmatrix} W^{EE} & W^{EI} \\ W^{IE} & W^{II} \end{pmatrix}$$

The  $N$  by  $N$  submatrices  $W^{EE}$  and  $W^{IE}$  are real and positive, whereas  $W^{EI}$  and  $W^{II}$  are real and negative, because of Dale's rule. Each sub-matrix  $W^{XY}$  is a translation invariant matrix. A translation invariant matrix is a matrix such that  $W_{(i,j)}^{XY} = w^{XY}(|j - i|)$ ,  $w^{XY}$  being an  $N$ -periodic function (see figure 3.16 A).

Murphy and Miller explain that for the family of matrices  $W$ , there is

a way of rotating  $W$  into a basis -called the Fourier basis- in which it is 2 by 2 block diagonal and in which it is straightforward to compute the Schur decomposition (see figure 3.16 B). Proceeding to do this, they realize that every connectivity matrix  $W$  of dimension  $2N$  by  $2N$  can be seen in the Schur basis as a collection of  $N$  feed-forward chains of length one (see figure 3.16 C,D,E). As we recall in the appendix 3.10.9, independently of the profiles, all translation invariant matrices of dimension  $N$  have the same eigenvectors  $\{e_1, e_2, \dots, e_N\}$ . The  $2N$  modes that compose the Schur decomposition, with  $N$  feed forward chains, are obtained concatenating the eigenvectors:

$$\left\{ \begin{pmatrix} e_1 \\ e_1 \end{pmatrix}, \begin{pmatrix} e_1 \\ -e_1 \end{pmatrix}, \begin{pmatrix} e_2 \\ e_2 \end{pmatrix}, \begin{pmatrix} e_2 \\ -e_2 \end{pmatrix}, \dots, \begin{pmatrix} e_N \\ e_N \end{pmatrix}, \begin{pmatrix} e_N \\ -e_N \end{pmatrix} \right\}$$

The eigenvectors are linked with the spatial Fourier decomposition of  $W$ :  $e_1$  is the zero-frequency (or DC component), whereas  $e_2$  is the first Fourier mode and its components are given by a cosine  $e_2|_j = \cos(\frac{2\pi j}{N})$ . The first feed-forward chain represents then the famous **balanced** amplification -from the difference to uniform pattern highlighted in the Murphy and Miller [105]. The other modes are called minus to plus modes. This decomposition is very interesting because it makes a direct link between the connectivity profile and the dynamical interpretation of the network. One trivial example is the case of  $W$  is when each submatrix is constant, so the spatial Fourier decomposition of the connectivity profile has only one DC component: the Schur decomposition is then reduced to only one feed-forward chain that goes from the difference to the uniform. Therefore, Murphy and Miller conjecture that if each submatrix has both strong excitation and inhibition characteristic of the balanced regime, and a biologically realistic fall of the connectivity strength with the distance, therefore these networks should have large balanced amplification (i.e. large variance in the uniform mode).

This argument is not asserted, the authors are not claiming that in all cases the variance of the uniform mode will be exactly higher than the variance of the others, but that the variance of the uniform will be very high. The purpose of our study (see appendix 3.10.9) was to study if we can find a connectivity matrix in the family of  $W$  such that the second feed forward chain has more variance than the first one. This would translate into competition.

We understood that one fundamental property of the translation invari-



ant connectivity matrices, is that given to their phase invariance, a cosine at a given frequency (example  $e_2|_j = \cos(\frac{2\pi j}{N})$ ) is an eigenvector of this matrix but that a sine ( $e_3|_j = \sin(\frac{2\pi j}{N})$ ) is also eigenvector of this matrix. Apart from the DC mode, for every frequency the eigenvectors come by pairs of cosines and sines, and have a same associated eigenvalue.

$$\begin{cases} \text{eigenvectors} : \{e_1, e_2, e_3, e_4, \dots, e_N\} \\ \text{eigenvalues} : \{\lambda_0, \lambda_1, \lambda_1, \lambda_2, \lambda_2, \dots, \lambda_{\frac{N}{2}}\} \end{cases}$$

This fact implies that the actual feedforward links and negative feedbacks of pairs of successive feedforward chains are equal (see figure 3.16 E and appendix 3.10.9). Therefore, if there is a parameter regime in which the variance of  $+_1$  is higher than the variance of U, the variance of  $+_2$  will also be higher than the variance of U. This means that if there is amplification, the dimension of the competitive subspace will be a multiple of two, and that is not what we observe in the data. In the appendix 3.10.9, we do all the mathematical demonstration of these claims and we show a particular example in which we can have this kind of competitive amplification.

Thus, if a translationally invariant connectivity was at the heart of the variability that we observe during cortical desynchronisation, we would expect noise to excite all phases of, at least, one particular Fourier mode, giving rise to a subspace of dimension two, and therefore, to two components with amplified variance in the PCA decomposition. The fact that we only observe one amplified dimension suggest to us that an approximately translation invariant connectivity is unlikely to underlie our observations.

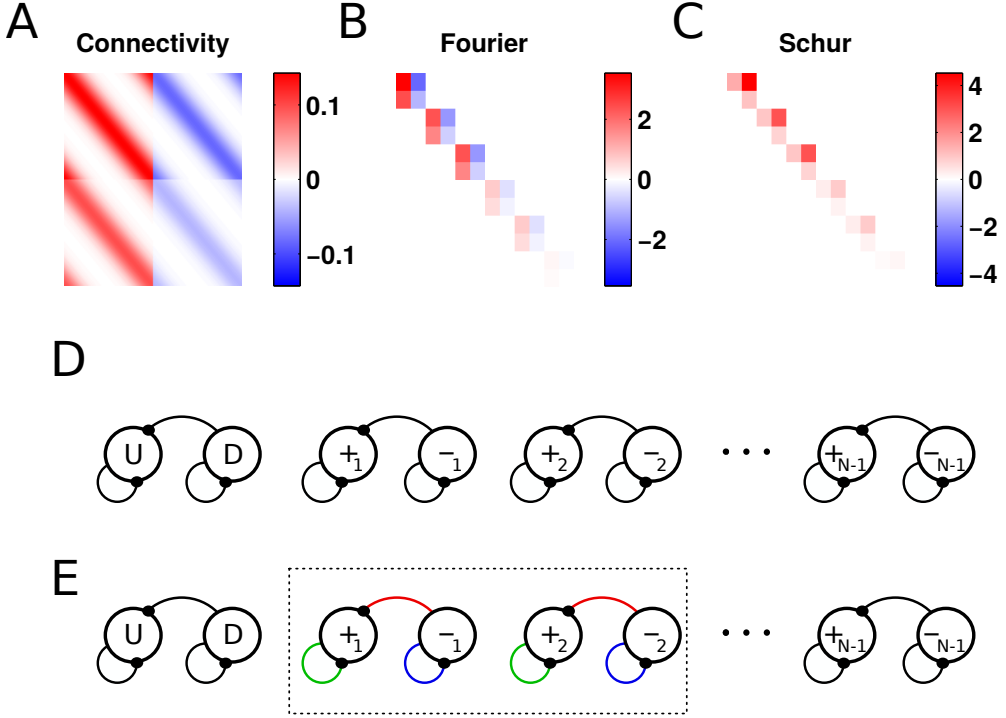


Figure 3.16: *A: translation invariant connectivity matrix. B: connectivity matrix in the Fourier basis (zoom of the first six modes). C: connectivity matrix in the Schur basis. D: Representation of the connectivity matrix in the Schur basis. E: connectivity in the Schur basis, noticing that although the modes are different,  $+_1$  and  $+_2$ , their associated eigenvalues are the same and therefore the negative feedbacks and the feedforward links of pairs of feedforward chains, are equal.*

## 3.8 Summary and discussion

### 3.8.1 Summary

In this chapter we developed the assumption that the connectivity constrains the spontaneous activity. If the dynamics can be simulated like a linear dynamical system, and if the dynamics are dominated by one pattern of activity, then the most amplified pattern of activity is quite aligned with the Schur vector of the connectivity matrix that has more variance.

We studied which kind of circuit patterns could generate competitive activity. We first went over the randomly connected network, and shown that it doesn't generate competitive amplification, so that we should add more structure to the connectivity matrices.

We first examined normal competitive amplification, which is a known mechanism that uses normal amplification, and then we introduced a variant of this basic circuit motif called transient competitive amplification. TCA uses a different type of computational mechanism (called non-normal amplification) to amplify patterns of activity.

Among NCA and TCA, TCA is the model that is more consistent with the data, and make strong predictions linking the variance of the two populations with the way they are correlated in time.

#### 3.8.2 Relation with previous work

Zagha et al [166] published in 2005 a similar study as ours in which they analyse data and do modelling, using three population model symmetrically connected. As we, they have recordings from layer 5 of rodent's cortex. However they record in motor cortex of awake behaving mice, while we record in auditory and somato-sensory cortex of urethane anesthetised rats. In their recordings, they discover negatively correlated activity between two populations of neurons. In their case, the population activity is very tightly linked to behaviour: one of the two populations is systematically enhanced -and the other systematically suppressed- after a whisker is touched, in anticipation of a correct lick of the mouse. Unlike in our recordings, the competitive activity is not alternating, but more categorical, "winner-take all" like, considering that in their case the same population always wins.

To model this process, they use a symmetric two excitatory population network that compete through mutual inhibition, but in the non-linear regime, as in Wang's papers [155], [160]. In order to create the asymmetry in the activity, they injected a constant external current to the suppressed population (equal to 20) and a brief strong current to the enhanced population of varying amplitude (equal at most to 210). If the pulse was strong enough, then the enhanced population stays active, but otherwise, falls back to baseline.

We then could ask ourselves whether we could generate asymmetry in the competition in the way Zagha et al [166] did, by increasing the variance

of the external noise in an NCA model to one of the populations. Even in parameter regimes that are not realistic, i.e for a noise injected to the first population having a standard deviation 100000 higher than the noise injected to the second population, the variance of the first population to the second remains of the order of 10/6, because the population E1 transfers variance to E2. The PC1 becomes less aligned with the uniform, like in the data. On what respects the cross-correlogram between E1 and E2, it is almost perfectly symmetric, and the small tiny positive lag that shows doesn't increases with the variance of the external noise. Therefore, the TCA model of asymmetric connectivity is more parsimonious, than an NCA with huge asymmetry in the external input.

An interesting approach, that we have in common with Zagha et al [166], is to try to deduce the circuit given the observed data. They propose a simple approach with is based on Granger causality, in order to disentangle connectivity scenarios, from the simple anti-correlated external inputs to a model like ours, with external uncorrelated inputs with mutual inhibition. With their data, they show that the Granger causality auto-regression analysis improves the prediction in time of one of the two populations, by taking into account in the regression the activity of the other population. We followed the same approach, an using the same toolbox they cite in the paper, taking the firing rate of one of the populations and seeing whether it predicts significantly better the other population. In our hands the result was not significant, which may indicate that they are not "Granger causally" related or alternatively that the method fails at picking that correlation. It would be however interesting to see the result of more elaborated methods like embedding.

Also, our study has similarities with a study from Litwin Kumar and Doiron [88]. The aim of their study is to observe the effect of subgroup clustering on the dynamics of the network: they observe that even for modest clustering, slow switching dynamics appeared: only one of the several clusters of the network is active at the time. As they use spiking networks, the mechanism underlying this process is not totally transparent, but seems to be similar to the NCA network: as the clustering index increases, when one cluster becomes active, it inhibits all the others clusters through the interaction with the inhibitory group. Schaub et al [132] used a rate model to explain the switching dynamics by Litwin Kumar and Doiron [88], and showed indeed that the normal amplification is involved in the slow dynam-

ics. However, they didn't considered the case in which there is non-normal amplification that could also explain the amplification.

#### 3.8.3 Functional relevance of the asymmetry

One of the key aspects of the TCA model is the asymmetry in the connectivity, which leads to an asymmetry in the dynamics. In the paper by Zagha et al. [166], because of the asymmetric input, in the end they also have an asymmetric circuit. Their hypothesis, is that the enhanced population feeds an integrator followed by a threshold, which triggers the licking behaviour. This hypothesis is in agreement with physiology, because in layer 5 precisely, the intra-encephalic neurons project to sub-cerebral projection neurons, and the reciprocal connection is very dim (see figure 1.1). The sub-cerebral projection neurons project to the brain stem and the spinal cord.

One hypothesis is then is that the asymmetry in the connectivity is the landmark of a functional difference, for example that the neurons from the E1 population are the output of the circuit. The other components of the circuit might be there to compete and feed this population. In our case, the competitive dynamics in spontaneous activity are not the same as the more discrete dynamics of Zagha et al., they are more analogical. One of the functions of this circuit might then be to integrate the sensory stimulus that arrives in a short window of time, and integrate it in the form of an oscillatory signal, whose instant amplitude depends precisely of the instantaneous intensity of the stimulus, like an amplitude modulated (AM) signal.

Another hypothesis about the role of the asymmetry, more in line with the null space hypothesis of evoked activity, presented in 2.3.4.5, is that the asymmetry in the connectivity favours that the amplitude of the firing in E1 be higher than the amplitude of the firing in E2. This might help orienting and confining the trajectories in the neural space to precisely the portion of space in which the firing rate of E1 is higher than the firing of E2 ( $E1 \geq E2$ ) (see figure 2.25). Given that the orthogonal complement of a subspace of dimension 1 is  $N-1$ , constraining the trajectories in such a high dimensional subspace would be helpful for decoding the different trajectories as a function of the different inputs.

### 3.8.4 On the specificity of model predictions

As we saw in the previous section, the model seems to reproduce both the qualitative aspects and to a good extent the quantitative aspects of the competitive amplification.

One of the untested predictions of the model concerns the role of the inhibitory neurons, mentioned in section 3.4.3. The TCA model, as we saw in the previous chapter, states that there is a small difference in the way each excitatory group is correlated with the inhibitory cells. However, as we saw in the computations and as we see in the simulations, the effect might be small, and depend on specific network parameters.

Peter Barthó, who recorded the data from primary somato-sensory cortex, applied an algorithm based on the spike shape that he had previously developed [8] in order to distinguish between excitatory and inhibitory neurons. In our data, the problem is that the number of identified putative inhibitory neurons is small (less than 15%, so around 9 neurons for a recording of 60 neurons), so we don't really test this prediction, but we are eager to do it with bigger simultaneous recordings.

Taken individually, some of the predictions made by NCA -the alternative model to TCA- might seem not so strong. Why would the first  $PC$  be exactly orthogonal to the uniform vector? Or, why would the variance of  $E_1$  be exactly similar to the variance of  $E_2$ ? Also, one might think that there are many possible ways of generating a lag in the cross-correlogram between  $E_1$  and  $E_2$ .

The advantage of the proposed model is that it is very parsimonious, so it states that all of these predictions are linked together and that they depend on one single parameter. For example, we showed in the previous chapter that we can label individual neurons based on their spiking power frequency, so that we blindly assign them as belonging to  $E_1$  or to  $E_2$ . Having put a label on each population, we can observe whether this group has a higher variance than  $E_2$ , and whether the cross-correlogram  $ccg(E_1, E_2)_{(\tau)}$  is systematically asymmetric with a positive lag.

### 3.8.5 Model fitting

The main achievement of the model is to provide an unified framework that explains a series of diverse measurements. In spite of this, one natural question we might ask is, can we fit this model to each experiment and obtain

a characterisation of the proportion of normal and non-normal amplification observed in each experiment ?

One suitable attribute of a model is that it doesn't require fine tuning for displaying the modelled phenomenon , so that the rough qualitative characteristics that the model replicates happen on a wide range of parameters. Yet, another suitable feature of a model is that the parameters have orthogonal effects on the outputs. Although both statements are not exclusive, they are rarely met in models with many parameters, so that one observation could be explained by many combinations of parameters.

In our model, a simple example of such redundancy is in the figure 3.13. In this figure, we observe that the noise on the connectivity adds variance to the deterministic modes, so that the crosscorrelograms are bigger. It is perfectly plausible that we can replicate those crosscorrelograms by using a model without noise in the connectivity and increasing the level of external noise, or by increasing the Schur coefficients in the same proportion, in order to increase the variance and the time scale of the modes.

Even if there is some redundancy, we could try to find all the solutions to compare them. Putting aside the five Schur coefficients over which we would do the search, there are some "hidden" parameters that have an influence on the result: the target spectral radius, the shape of the competitive vector, the probability of connection  $p$ . From the experimental size, we can use the first principal component to orient the shape of the competitive pattern as well as anatomical studies to refine the choice of the proportion of connected neurons.

Nevertheless, there are maybe some conceptual bottlenecks that should be solved before comparing quantitatively in-vivo parallel recordings with a network model or in particular with TCA.

On one side, from a completely theoretical perspective, we should have a clear view of whether we can completely reconstruct a high dimensional linear system when we sample randomly certain dimensions from it, like during recordings. It would be good to know if only based on measures, sampled from artificial spiking networks, it is possible to have an idea about the dimensionality of the dynamical system that generated the data. Recent work by Williamson et al [158] seems to go on this direction.

Also, we know from last chapter that the competition decreases with the distance. On one hand this fact points in the right direction, because it shows that the competition is local, and this means that we can model

it with a relatively small network. However, it also suggest that we should incorporate constraints on the connectivity like a decrease with the distance (see [126]), or boundary conditions. From a modelling perspective, we can suppose that we are sampling from an extended dynamical system which has local properties like competition, and this could lead to refine conceptually what we call external uncorrelated input  $\vec{\xi}$ , which might simply be activity from more distant neurons.

#### 3.8.6 More experiments to refine the predictions ?

In order to test properly the model with regard to the role of inhibitory cells in the competition, and maybe distinguishing TCA from UCD (see section 3.7.1), it would be interesting of doing more experiments identifying directly inhibitory cells by infecting them with channel-rhodopsin and activating them specifically with light.

When doing more experiments to study the spatial aspect of the competition, it should be possible to use calcium imaging to have a better sense of the entire circuit and of the spatial arrangement of the cells, even if negative correlations are harder to detect with this technique. The way we constructed the connectivity matrix shows that to obtain low dimensional dynamics it is not necessary to have a low dimensional connectivity: it suffices that the connectivity is partially structured and partially random, and that they interact in a particular way. One interesting experiment would be to do as Ko et al [82], to first observe the competition in-vivo and then probe in-vitro how strong the neurons are connected between themselves. When sorting the neurons by loading -and not by distance, because we saw in the previous paragraph that the competition happens locally-, we think that we would see an underlying slowly varying spatial continuity of the connection strengths (figure 3.11 D).

#### 3.8.7 Other possible experiments to expand our understanding about the role of competition

As we already said in the introduction, correlating in-vivo recordings with sensory stimulation in awake animals would be a laborious but important step for understanding of the role of the desynchronised state in sensory processing, and the way attention modulates it. Now that we understand



that the asynchronous state is competitive, we can observe what happens with both groups of neurons during sensory stimulation.

Also, it would be interesting to see if we can achieve competition in slices of layer 5, in a similar experiment as the one done by Sanchez-Vivez and McCormick [131], in which they observed locally generated synchronous activity. This would allow to add different neurotransmitters on the preparation and observe the effects on the competition.

## 3.9 Methods

### 3.9.1 Balanced randomly connected network simulation

We simulated 10 minutes of activity of the balanced network using Brian, a simulator for spiking neural networks [53]. All the differential equations concerning the membrane potentials, the synaptic currents, and the conductances are explained in detail in the supplementary information of [123].

We reproduce here the parameters used for the simulation of figures 3.4 and 3.5, as in ([123]):  $N_E = 4000$ : number of excitatory neurons.  $N_I = 1000$ : number of inhibitory neurons.  $N_x = 1000$ : number of external Poisson neurons.  $\mu_{ext} = 2.5 \text{ Hz}$ : firing rate of external Poisson neurons.  $p = 0.2$ : connection probability for the feed-forward and the recurrent connections.  $C_m = 0.25 \text{ nF}$ : membrane capacitance.  $g_L = 16.6 \text{ nS}$ : leak conductance.  $\tau_r = 1 \text{ ms}$ : rise time of the unitary conductance change.  $\tau_d = 5 \text{ ms}$ : decay time of the unitary conductance change.  $\theta = -50 \text{ mV}$ : spike threshold.  $V_L = -70 \text{ mV}$ : resting membrane potential.  $V_R = -60 \text{ mV}$ : reset value.  $\tau_{ref}^E = 2 \text{ ms}$ : refractory period excitatory.  $\tau_{ref}^I = 1 \text{ ms}$ : refractory period inhibitory cells. Mean synaptic conductances:  $g_{EE} = 2.4 \text{ nS}$ ,  $g_{EI} = 4.8 \text{ nS}$ ,  $g_{IE} = g_{II} = 40 \text{ nS}$ ,  $g_{IX} = g_{IEX} = 5.4 \text{ nS}$ . Standard deviation of synaptic conductances:  $\sigma_{XY}^g = 0.5 g_{XY}$ . In order to compute the eigenspectrum, we applied the same procedure that we used for real data (see methods of the previous chapter).  $t_c = 0.1 \text{ s}$  and  $N_{jitt} = 10$ .

### 3.9.2 The Lyapunov equation

We introduce the **Lyapunov equation**, which we will use extensively in the rest of this study. Given a system described by an Ornstein-Uhlenbeck process :  $\dot{X} = (J - I)X + B\xi$ , the Lyapunov equation [47] assures that, given the system drift  $J - I$ , and its diffusion term  $B$ , the covariance matrix  $C$  is obtained solving the equation:

$$(J - I)C + C(J - I)^t = -BB^t$$

This equation can not be solved analytically except for very low dimensional cases. It is solved using a recursive routine that first rotates the

connectivity matrix to a Schur basis (to obtain a upper-triangular connectivity matrix), and then solves recursively the variances/covariances starting from below. Once all the correlations are solved, the routine rotates back to the canonical basis. The complexity of this algorithm is of the order of  $N^2$ , where  $N$  is the dimension of the system. In Matlab, the routine is called "lyap".

#### 3.9.3 Mean trajectory and cross-correlograms for linear dynamical systems

For an Ornstein-Uhlenbeck processes we can compute the mean trajectory and the cross-correlogram [47]. The cross-correlogram measures the degree with which two random variables fire on average around their mean, with a certain temporal lag between the variables.

The mean trajectory, given a series of initialisations is:

$$\langle X(t) \rangle = e^{(J-I)t} \langle X_{(t=0)} \rangle$$

Similarly, the cross-correlogram is:

$$\langle X_{(t)} X_{(t+\tau)}^t \rangle = \begin{cases} e^{(J-I)\tau} C, & \tau > 0 \\ C \cdot e^{-(J-I)^t \tau}, & \tau < 0 \end{cases} \quad (3.1)$$

Considering the propagator matrix  $e^{(J-I)\tau}$ , we note that the same mechanism that may generate transient amplification on the mean trajectories for non-normal matrices (see figure 3.2), may also generate transient amplification on the cross-correlograms. By definition, the autocorrelograms are symmetric, but the cross-correlograms doesn't have to be.

#### 3.9.4 Cross-correlograms of spike trains

Let's define the spike train of cell  $i$  as :

$$\rho_i(t) = \sum_{m=1}^{n_i} \delta(t - t_m)$$

The mean firing rate can be written as :  $r_i = \frac{1}{T} \int_0^T d\tau \cdot \rho(\tau)$

The crosscorrelogram of the spike trains of the cells  $p$  and  $q$  can be defined as [33]:

$$Q_{\rho_p \rho_q}(\tau) = \frac{1}{T} \int_0^T dt. (\rho_p(t) - r_p)(\rho_q(t+\tau) - r_q) = \frac{1}{T} \int_0^T dt. \rho_p(t) \cdot \rho_q(t+\tau) - r_p \cdot r_q$$

When

$$\tau \rightarrow T, \frac{1}{T} \int_0^T dt. \rho_p(t) \cdot \rho_q(t+\tau) - r_p \cdot r_q \rightarrow 0$$

The discretized version of the crosscorrelogram is :

$$\begin{aligned} Q^*(m) &= \frac{1}{\Delta t} \int_{(m-\frac{1}{2})\Delta t}^{(m+\frac{1}{2})\Delta t} Q_{\rho_p \rho_q}(\tau) \cdot d\tau \\ &= \frac{1}{T} \frac{1}{\Delta t} \int_{(m-\frac{1}{2})\Delta t}^{(m+\frac{1}{2})\Delta t} \int_0^T dt. d\tau. (\rho_p(t) \cdot \rho_q(t+\tau)) - r_p \cdot r_q, \quad m \in N \end{aligned}$$

We use the following normalized version of the crosscorrelogram:

$$Q_{scale}^*(m) = \frac{\frac{1}{T} \frac{1}{\Delta t} \int_{(m-\frac{1}{2})\Delta t}^{(m+\frac{1}{2})\Delta t} \int_0^T dt. d\tau. (\rho_p(t) \cdot \rho_q(t+\tau))}{r_p \cdot r_q}$$

Therefore,  $Q_{scale}^*(m) \rightarrow 1$  when  $m \rightarrow +\infty$ .

For each experiment in figure 3.10, we used a time window of  $t_c = 0.1$  s. For each pair of cells, we computed two cross-correlograms: one on the raw data, and another on a surrogate version of the data, in which we randomly reassigned the spikes inside fixed jitter windows of length 10 time windows. We then subtracted the two cross-correlograms. Finally, once we had computed all the ccg, we average them accross pairs -omitting the autocorrelograms.

This procedure is related to the way we pre-process the data, and guarantees that we put an upper bound on the time scale at which the cross-correlations are happening (see methods 2.4.2 in chapter 2).

Instead of averaging the ccgs, we could have also computed a weighted sum loading each ccg by the product of loadings of each cell. In practice, with this other method, the results are qualitatively similar to the ones presented in figure 3.10.

## 3.10 Appendix

### 3.10.1 Two dimensional linear excitatory and inhibitory network

In this paragraph, we deduce the Schur representation from the anatomical representation, and we also do the converse -from the Schur domain to the canonical basis. Though this exercise we introduce the vocabulary, the notation and justify subsequent results.

In the canonical basis, a network composed of an excitatory population and an inhibitory one  $X = \begin{pmatrix} x_E(t) \\ x_I(t) \end{pmatrix}$ , obeys the differential equation  $\dot{X} = -X + JX + \xi$  and has the following connectivity matrix:

$$J = \begin{pmatrix} J_E & -\Delta - J_E \\ J_I & -\Delta - J_I \end{pmatrix}$$

We can see that this matrix has only three degrees of freedom  $J_E$ ,  $J_I$  and  $\Delta$ , instead of four. This is a modelling choice, further justified by Hennequin (see Appendix of [64]). In this configuration, the matrix follows the constraint that the **uniform** vector

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

is an eigenvector, whose associated eigenvalue is  $\lambda_1 = -\Delta \ll 0$ . This means that a uniform perturbation of the activity will be suppressed fast. Also, we are in the so called **inhibition-dominated regime**: both the total amount of excitation and of inhibition are big (order 1) and the total amount of inhibition exceeds the amount of excitation. The total amount of excitation and inhibition received by each population is constant:  $J_E + (-\Delta - J_E) = J_I + (-\Delta - J_I) = -\Delta$ .

The second eigenvector and its associated eigenvalue are:

$$v_2 = \frac{1}{\sqrt{1+k^2}} \begin{pmatrix} k \\ 1 \end{pmatrix}, \quad k = \frac{\Delta + J_E}{J_I}, \quad \lambda_2 = J_E - J_I$$

We then choose as first Schur vector  $u_1 = v_1$  and compute the second Schur vector through Gram-Schmidt:

$$u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

This vector is called the **difference** vector. We will designate it by the letter D. Applying the change of basis formula, we deduce that in the Schur basis, the functional representation T of the connectivity is:

$$T = \begin{pmatrix} \lambda_U & t \\ 0 & \lambda_D \end{pmatrix}, \quad \lambda_U = \lambda_1 = -\Delta, \quad \lambda_D = \lambda_2 = J_E - J_I, \quad t = J_E + J_I + \Delta$$

The basis rotation  $X = PY$  transforms the differential equation  $\dot{X} = -X + JX + \xi$  into  $\dot{Y} = -Y + TY + P\xi$ . Given that the external noise received by each population is uncorrelated  $\langle \xi_E \cdot \xi_I \rangle = 0$ ,  $P\xi$  is equivalent to  $\xi$  because P is unitary and  $B = I$ .

We can do the same exercise the other way around: starting from the Schur basis and deducing the connectivity matrix.

$$J = P \cdot \begin{pmatrix} \lambda_U & t \\ 0 & \lambda_D \end{pmatrix} P^t, \quad P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$J = \frac{1}{2} \begin{pmatrix} \lambda_U + \lambda_D + t & \lambda_U - \lambda_D - t \\ \lambda_U - \lambda_D + t & \lambda_U + \lambda_D - t \end{pmatrix}$$

$$\begin{cases} J_E &= \frac{\lambda_U + \lambda_D + t}{2} \\ J_I &= \frac{\lambda_U - \lambda_D + t}{2} \\ \Delta &= -\lambda_U \end{cases}$$

### 3.10.2 Study of the correlation in the 2D linear network

Given that we just built the connectivity matrix in two different ways (see previous paragraph), we now have the tools to demonstrate that the two excitatory and inhibitory populations tend to be positively correlated.

We could try to solve directly the Lyapunov equation in the neural space, and show that the term  $C_{1,2} = C_{EI}$  tends to be positive. However, the equations are very cumbersome. Instead of that, we will make the demonstration in two steps:

a) Given the connectivity matrix in the Schur space, show that the uniform mode has more variance than the difference mode.

b) Given the variances and covariances in the Schur space, show how does the covariance  $C_{EI}$  depends on the covariances of the uniform and difference modes.

**Step a:** Lyapunov equation in the Schur space.

Lets note  $C_s$  the covariance matrix and  $T$  the connectivity matrix in the Schur space. The Lyapunov equation is:

$$(E) : (T-I)C_s + C_s(T-I)^t = -I, \quad T = \begin{pmatrix} \lambda_U & t \\ 0 & \lambda_D \end{pmatrix}, \quad C_s = \begin{pmatrix} c_U & c_{ud} \\ c_{ud} & c_D \end{pmatrix}$$

Developing the Lyapunov equation gives:

$$(E) \iff \begin{cases} 2c_U(\lambda_U - 1) + 2c_{ud}t = -1 \\ c_{ud}(\lambda_U + \lambda_D - 2) + t.c_D = 0 \\ 2c_D(\lambda_D - 1) = -1 \end{cases}$$

$$(E) \iff \begin{cases} c_U = \frac{-1}{2(\lambda_U-1)} + c_D \frac{t^2}{(\lambda_U-1)(\lambda_U+\lambda_D-2)} \\ c_{ud} = -\frac{t.c_D}{\lambda_U+\lambda_D-2} \\ c_D = \frac{-1}{2(\lambda_D-1)} \end{cases}$$

We then concentrate on the first equation ( $c_U = \frac{-1}{2(\lambda_U-1)} + c_D \frac{t^2}{(\lambda_U-1)(\lambda_U+\lambda_D-2)}$ ), to show that  $c_U > c_D$ , we just need to show that  $\frac{t^2}{(\lambda_U-1)(\lambda_U+\lambda_D-2)} > 1$ . This term can be rewritten as:

$$\frac{t^2}{(\lambda_U - 1)(\lambda_U + \lambda_D - 2)} = \frac{t^2}{(t + (1 - J_E) - J_I)(t + 2(1 - J_E))}$$

Then, it suffices that  $J_E > 1$ , to get  $c_U > c_D$ , because the numerator is bigger than the denominator ( $t \gg 1$  and  $t - \epsilon_1 < t$  and  $t - \epsilon_2 < t$ ).

This condition is then true in the usual range of the parameters, specially in the case in the inhibition dominated regime, in which the excitation and the inhibition are thought to be big.

#### Step b: Change of basis.

If we write the Lyapunov equation in both the Schur basis and in the canonical basis, we can deduce that there is a change of basis formula for the covariance:

$$C = P.C_s.P^t, \quad P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad C_s = \begin{pmatrix} c_U & c_{ud} \\ c_{ud} & c_D \end{pmatrix}$$

We deduce easily the covariance in the canonical basis:

$$C = \frac{1}{2} \begin{pmatrix} c_U + c_D + 2c_{ud} & c_U - c_D \\ c_U - c_D & c_U + c_D - 2c_{ud} \end{pmatrix}$$

The covariance between the excitation and the inhibition is then the difference between the variances of the uniform mode and the variance of the difference mode. As we just shown, this difference is positive.

### 3.10.3 Rotating the covariance matrix from the Schur to the canonical basis

Starting from the differential equation of the system, we can apply an axis rotation from the canonical to the Schur basis making the change of variable  $PY = X$ .

$$\dot{X} = -X + JX + \xi \text{ and } \dot{Y} = -Y + TY + P^t.\xi$$

If we pose the Lyapunov equation in the Schur basis, we obtain the covariance matrix  $C_s$  between the different Schur modes:

$$(T - I)C_s + C_s(T - I)^t = -I$$

Comparing the Lyapunov equation in both basis, and recalling that  $J = PTP'$ , we obtain the change of basis formula for the covariances in the canonical basis:

$$P.C_s.P^t = C$$



$$\left\{ \begin{array}{l} c_{E_1 E_1} = \frac{c_{cc}}{2} + \frac{c_{uu}}{3} + \frac{c_{dd}}{6} + \frac{\sqrt{6}c_{cu}}{3} + \frac{\sqrt{3}c_{cd}}{3} + \frac{\sqrt{2}c_{ud}}{3} \\ c_{E_1 E_2} = \frac{c_{uu}}{3} - \frac{c_{cc}}{2} + \frac{c_{dd}}{6} + \frac{\sqrt{2}c_{ud}}{3} \\ c_{E_1 I} = \frac{c_{uu}}{3} - \frac{c_{dd}}{3} + \frac{\sqrt{6}c_{cu}}{6} - \frac{\sqrt{3}c_{cd}}{3} - \frac{\sqrt{2}c_{ud}}{6} \\ c_{E_2 E_2} = \frac{c_{cc}}{2} + \frac{c_{uu}}{3} + \frac{c_{dd}}{6} - \frac{\sqrt{6}c_{cu}}{3} - \frac{\sqrt{3}c_{cd}}{3} + \frac{\sqrt{2}c_{ud}}{3} \\ c_{E_2 I} = \frac{c_{uu}}{3} - \frac{c_{dd}}{3} - \frac{\sqrt{6}c_{cu}}{6} + \frac{\sqrt{3}c_{cd}}{3} - \frac{\sqrt{2}c_{ud}}{6} \\ c_{II} = \frac{c_{uu}}{3} + \frac{2c_{dd}}{3} - \frac{2\sqrt{2}c_{ud}}{3} \end{array} \right.$$

### 3.10.4 Spectral decomposition

The aim of this appendix is demonstrate a simple restatement of the change of basis formula  $J = PTP^t$  as a sum of external products known as the **spectral decomposition**. This formula is useful for understanding how the parameters of the Schur matrix T and the Schur vectors shape the connectivity matrix in the canonical basis.

J is a connectivity matrix. J admits a Schur decomposition  $J = PTP'$ , that allows to pass from a recurrently connected network, to an upper-triangular feed-forward network. P is an unitary matrix whose columns are the vectors from the Schur basis, that is obtained orthogonalizing the eigenvectors of J.

$$I = PP^t = P^t.P$$

$$P = \begin{pmatrix} \vec{u}_1 & \vec{u}_2 & \dots & \dots & \vec{u}_N \end{pmatrix}$$

First, lets use the equality :  $J = P.T.P^t$  to see how the Schur vectors  $\vec{u}_i$  are expressed in terms of elements of the the Schur matrix:

$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1N} \\ 0 & t_{22} & t_{23} & \dots & t_{2N} \\ 0 & 0 & t_{33} & \dots & \dots \\ 0 & 0 & \dots & \dots & t_{NN} \end{pmatrix}, \text{ with } t_{ii} = \lambda_i.$$

$$J.\vec{u}_i = P.T.P^t.\vec{u}_i = P.T. \begin{pmatrix} 0 \\ \dots \\ 0 \\ 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} = P. \begin{pmatrix} t_{1i} \\ t_{2i} \\ \dots \\ t_{ii} \\ 0 \\ \dots \\ 0 \end{pmatrix} = \sum_{j=1}^i t_{ji} \vec{u}_j$$

$$J = J.P.P^t = J. \begin{pmatrix} \vec{u}_1 & \vec{u}_2 & \dots & \dots & \vec{u}_N \end{pmatrix} . \begin{pmatrix} \vec{u}_1^t \\ \vec{u}_2^t \\ \vec{u}_3^t \\ \dots \\ \vec{u}_N^t \end{pmatrix} = J. \sum_{i=1}^N \vec{u}_i \vec{u}_i^t =$$

$$\sum_{i=1}^N (J.\vec{u}_i) . \vec{u}_i^t = \sum_{i=1}^N (\lambda_i . \vec{u}_i + \sum_{j=1}^{i-1} T_{ji} \vec{u}_j) . \vec{u}_i^t$$

Then:

$$J = \sum_{i=1}^N \lambda_i . \vec{u}_i . \vec{u}_i^t + \sum_{i=1}^N \sum_{j=1}^{i-1} T_{ji} \vec{u}_j . \vec{u}_i^t$$

#### 3.10.4.1 Corollary of the spectral decomposition: the difference to uniform block assures Dale's law

Murphy and Miller [105] showed that biologically inspired networks that follow Dale's law have a strong feed-forward link in the Schur domain, from the difference to the uniform mode. Conversely, when we build our connectivity matrix starting from the Schur basis, the spectral decomposition shows us that precisely the link between the difference to the uniform mode is the one that has to be big enough so that the connectivity matrix remains Dale.

The spectral decomposition shows how the connectivity matrix results from the superposition of matrices composed by external products of Schur vectors. For example, in the case of the network composed of excitatory and inhibitory neurons:

$$J = P. \begin{pmatrix} \lambda_U & t \\ 0 & \lambda_D \end{pmatrix} P^t = \lambda_u . \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \lambda_d . \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + t \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$$

When we build the connectivity matrix starting from the connectivity matrix in the Schur space, the link  $t$ , that multiplies the matrix  $\frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} = \vec{U} . \vec{D}^t$  has to be big enough so that Dale is preserved.

#### 3.10.5 Computing the eigenvectors of a connectivity matrix

When we build the connectivity matrix starting from the Schur basis, one might want to know the eigenvectors. The recipe to do so is to re-multiply the connectivity basis by the Schur vectors:

$$J.u_k = \left( \sum_{i=1}^N \lambda_i \vec{u}_i \cdot \vec{u}_i^t + \sum_{i=1}^N \sum_{j=1}^{i-1} T_{ji} \vec{u}_j \cdot \vec{u}_i^t \right) . u_k, \quad k = 1, \dots, N$$

and then solve the system first identifying when possible the eigenvectors with the Schur vectors  $J.u_k = \lambda_k . u_k \implies u_k = v_k$  and otherwise posing that the eigenvectors should be a linear combination of the Schur vector, and have unitary norm.

We can do it for the connectivity matrix of TCA:  $T = \begin{pmatrix} \lambda_C & M & 0 \\ 0 & \lambda_U & V \\ 0 & 0 & \lambda_D \end{pmatrix}$ .

$$\begin{cases} J.u_1 &= \lambda_c . u_1 \\ J.u_2 &= \lambda_u . u_2 + M . u_1 \\ J.u_3 &= \lambda_d . u_3 + V . u_2 \end{cases}$$

We can see that the first eigenvector and eigenvalue is  $v_1 = u_1$  and  $\lambda_1 = \lambda_c$

Lets compute the next eigenvector, knowing that it will be a linear combination of  $u_1$  and  $u_2$ , and by definition, will have norm equal to one.

$$J.(\sqrt{1 - \beta^2} . u_2 + \beta . u_1) = \lambda_u . (\sqrt{1 - \beta^2} . u_2 + \frac{M\sqrt{1 - \beta^2} + \beta\lambda_c}{\lambda_u} . u_1)$$

$$\beta = \frac{M\sqrt{1 - \beta^2} + \beta\lambda_c}{\lambda_u} \implies \beta = \frac{M}{\sqrt{(\lambda_u - \lambda_c)^2 + M^2}}$$

We then conclude that the second eigenvector is  $v_2 = \sqrt{1 - \beta^2} . u_2 + \beta . u_1$ ,  $\beta = \frac{M}{\sqrt{(\lambda_u - \lambda_c)^2 + M^2}}$  and the second eigenvalue is:  $\lambda_2 = \lambda_u$ .

The third eigenvector and eigenvalue is, similarly, a linear combination of  $v_2$  and  $u_3$ :

$$v_3 = \sqrt{1 - \alpha^2} . v_2 + \alpha . u_3, \quad \alpha = \frac{V}{\sqrt{V^2 + (\lambda_d - \lambda_u)^2}}, \quad \lambda_3 = \lambda_d$$

### 3.10.6 Schur decomposition of a 2\*2 matrix

In this paragraph, we will present the Schur decomposition of a 2\*2 Dale matrix, adopting a convention in which all the coefficients are positive:

$$\begin{pmatrix} w_{EE} & -w_{EI} \\ w_{IE} & -w_{II} \end{pmatrix}$$

Lets consider two un-normalized eigenvectors of the kind:  $\begin{pmatrix} 1 \\ x_{\pm} \end{pmatrix}$

$$\begin{pmatrix} w_{EE} & -w_{EI} \\ w_{IE} & -w_{II} \end{pmatrix} \begin{pmatrix} 1 \\ x_{\pm} \end{pmatrix} = \begin{pmatrix} w_{EE} - x_{\pm}w_{EI} \\ w_{IE} - x_{\pm}w_{II} \end{pmatrix} = (w_{EE} - x_{\pm}w_{EI}) \begin{pmatrix} 1 \\ \frac{w_{IE} - x_{\pm}w_{II}}{w_{EE} - x_{\pm}w_{EI}} \end{pmatrix}$$

We can then solve the equation :

$$\frac{w_{IE} - x_{\pm}w_{II}}{w_{EE} - x_{\pm}w_{EI}} = x_{\pm} \iff \Delta = (w_{II} + w_{EE})^2 - 4w_{EI}w_{IE} > 0$$

$$\text{and } x_{\pm} = \frac{w_{II} + w_{EE} \pm \sqrt{\Delta}}{2w_{EI}}$$

Then the eigenvalues are:  $\lambda_{\pm} = w_{EE} - x_{\pm}w_{EI}$ . The normalized eigenvectors can then be written:

$$e_+ = \frac{1}{\sqrt{1+|x_+|^2}} \begin{pmatrix} 1 \\ x_+ \end{pmatrix} \text{ and } e_- = \frac{1}{\sqrt{1+|x_-|^2}} \begin{pmatrix} 1 \\ x_- \end{pmatrix}. \text{ These eigenvectors are not orthogonal, so lets establish a Schur basis } \{e_+, z\}, \text{ Where } z = \frac{e_- - (e_- \cdot e_+)e_+}{\sqrt{1 - |e_- \cdot e_+|^2}} = \frac{1}{\sqrt{1+|x_+|^2}} \begin{pmatrix} x_+ \\ -1 \end{pmatrix}$$

We can then rewrite  $e_- = z \cdot \sqrt{1 - |e_- \cdot e_+|^2} + (e_- \cdot e_+)e_+$ . Multiplying the matrix by z:

$$\begin{aligned}
 \begin{pmatrix} w_{EE} & -w_{EI} \\ w_{IE} & -w_{II} \end{pmatrix} z &= \begin{pmatrix} w_{EE} & -w_{EI} \\ w_{IE} & -w_{II} \end{pmatrix} \left( \frac{e_- - (e_- \cdot e_+)e_+}{\sqrt{1 - |e_- \cdot e_+|^2}} \right) = \frac{\lambda_- e_- - \lambda_+ (e_- \cdot e_+)e_+}{\sqrt{1 - |e_- \cdot e_+|^2}} \\
 &= \frac{\lambda_-}{\sqrt{1 - |e_- \cdot e_+|^2}} (z \cdot \sqrt{1 - |e_- \cdot e_+|^2} + (e_- \cdot e_+)e_+) - \frac{\lambda_+ (e_- \cdot e_+)e_+}{\sqrt{1 - |e_- \cdot e_+|^2}} \\
 &= \lambda_- z + \frac{(\lambda_- - \lambda_+)(e_- \cdot e_+)}{\sqrt{1 - |e_- \cdot e_+|^2}} e_+ \\
 &= \lambda_- z + t \cdot e_+
 \end{aligned}$$

In the Schur basis, this 2\*2 matrix can be expressed as:  $\begin{pmatrix} \lambda_+ & t \\ 0 & \lambda_- \end{pmatrix}$

#### Particular solution

The previous solution is very much simplified if we impose the constraint that the sum of the rows is constant:  $w_{EE} - w_{EI} = w_{IE} - w_{II}$ . This implies that  $x_{\pm} = \frac{w_{EE} + w_{EE} \pm |w_{EI} - w_{IE}|}{2w_{EI}}$ . Given that the absolute value function gives the same results in the two cases  $w_{EI} > w_{IE}$  or  $w_{EI} < w_{IE}$ , and that this equation has two solutions 1 and  $\frac{w_{IE}}{w_{EI}}$ , we can always choose  $x_+$  to be equal to 1.

We can then compute easily the eigenvectors of the Fourier decomposition and the feed-forward link.

$$\begin{cases} \lambda_+ &= w_{EE} - w_{EI} \\ \lambda_- &= w_{EE} - w_{IE} \\ t &= w_{EI} + w_{IE} \end{cases}$$

### 3.10.7 Adding the deterministic and the stochastic high dimensional connectivity matrices

In general when we pose the question of how does the eigenspectrum of the sum of two matrices A and B looks with respect to the eigenspectra of A and B, there is not a simple answer.

In the particular case in which A and B have the same eigenvectors, the eigenspectrum of the sum matrix  $A + B$  is the sum of the eigenspectra of A and B:

$$\begin{aligned} A + B &= P_A \cdot D_A \cdot P_A^{-1} + P_B \cdot D_B \cdot P_B^{-1} = P_A \cdot D_A \cdot P_A^{-1} + P_A \cdot D_B \cdot P_A^{-1} \\ &= P_A \cdot (D_A + D_B) \cdot P_A^{-1} = P \cdot (D_A + D_B) \cdot P^{-1} \end{aligned}$$

We now consider M to be the deterministic (low rank) part of a connectivity matrix and F to be the stochastic part.

**Lemma** (Generalisation of Rajan and Abbott [122]):

Define  $M \in \mathbb{R}^{N \times N}$  a real square matrix of rank k. Define also  $F \in \mathbb{R}^{N \times N}$ . If we impose k conditions on the matrix F:  $F \cdot v_{1,\dots,k} = 0$ , then M and F have the same eigenvectors, and M+F inherit k eigenvalues from M and  $N - k$  eigenvalues from F.

**Proof:**

If  $l$  is a left eigenvector of F:

$$\forall k, l.(F \cdot v_k) = l \cdot \vec{0} = 0 = \lambda_l l \cdot v_k$$

$$\text{If } \lambda_l \neq 0 \text{ then } l \cdot v_k = 0$$

M has k eigenvectors  $v_{1,\dots,k}$ . The image of M is equal to the vector space generated by linear combinations of the eigenvectors  $Im(M) = \langle v_1, v_2, \dots, v_k \rangle$ .

$$\forall v : l.(M \cdot v) = l.(\sum_{i=1}^k \alpha_i v_i) = (\sum_{i=1}^k \alpha_i l \cdot v_i) = 0 \iff l \cdot M = 0$$

Then

$$\boxed{l(M + F) = l \cdot F = \lambda_l \cdot l}$$

**Corollaries:**

**Corollary 1:**

Having imposed the above mentioned condition on F, the Schur decomposition of F+M is the sum of the Schur decomposition of F and the Schur decomposition of M (see figure 3.12):

$$T_{F+M} = T_F + T_M$$

**Corollary 2:**

$$\begin{cases} F.v_1 = 0 \\ F.v_2 = 0 \\ \dots \\ F.v_k = 0 \end{cases} \iff \begin{cases} F.u_1 = 0 \\ F.u_2 = 0 \\ \dots \\ F.u_k = 0 \end{cases}$$

Proof: Gram-Schmidt decomposition.

As the Schur vectors  $u$  are orthogonal between them, we can impose serially the condition  $F.u_i = 0$ ,  $i = 1, \dots, 3$ , by doing :

$$F(r, :) \leftarrow F(r, :) - (F(r, :).u_i).u_i$$

A more direct way to do it is to do:

$$F \leftarrow F.(I - \sum_{i=1}^k u_i.u_i')$$

**Corollary 3:**

Nil elements of the deterministic and of the stochastic part of the connectivity matrices, in the Schur basis.

- **Deterministic matrix:**

$$\forall (i, j) \notin 1, 3 \quad (T_M)_{ij} = 0$$

Proof:  $T_M \in \mathbb{R}^{N.N}$  is obtained by zero-padding  $T$  of dimension  $< 3.3 >$ , the low dimensional Schur representation of the TCA model.

- **Stochastic matrix:**

$$\forall (i, j) \in \{1, 2, 3\}, (T_F)_{ij} = 0.$$

Proof:  $\forall i, j : i > j, i, j \in \{1, 2, 3\} :$

$$\begin{aligned}(T_F)_{ij} &= (P'.F.P)_{ij} \\ &= P'.F.u_j \\ &= P'.0 = 0\end{aligned}$$

#### 3.10.7.1 Different types of noise that constitute the stochastic matrix

We can have many types of noise to with which to fill the matrix  $F$ . Here we will cover two types of noise:

##### a) Fully connected network with Gaussian noise

The stochastic component of the connectivity between each every pair of units is sampled independently from a Gaussian distribution of zero mean and standard deviation equal to  $\frac{R}{\sqrt{N}}$ .  $F_{i,j} \sim N(0, \frac{R}{\sqrt{N}})$ . Girko's circle law states that the eigenspectrum of  $F$  will be uniformly distributed inside a circle of radius  $R$ . After filling the matrix  $F$ , we impose the condition  $F \leftarrow F \cdot (I - \sum_{i=1}^3 u_i)$ . This operation doesn't change much the eigenspectrum, but nevertheless eliminates the outliers, the eigenvalues that are out of the circle and which could destabilize the system [122].

In the Gaussian distribution, the mean and the variance are two independent parameters, so we can control the spectral radius as we wish. However, in the case of Bernoulli noise, the mean (proportion of connections)  $p$  and the variance  $p(1-p)$  are linked, and this will imply some technical complications that we detail in what follows.

##### b) Sparsely connected network with Bernoulli noise

Bernoulli noise is the standard kind of noise that is used when studying recurrent networks [123], [64]. For every pair of neurons if a randomly generated



number is between zero and  $p$ , we establish a connection (the strength can vary depending on the identity of the pre-synaptic and of the post-synaptic neurons). If we sample the connections with this method, but keeping the actual value of the deterministic matrix  $M(i, j)$  when sampling the connection between neurons  $i$  and  $j$ , we are not going to maintain the eigenvalues and eigenvectors of  $M$ . This would happen for example, if we averaged across many sampled random matrices, but in that case we would lose the all or none character of the connections.

Using the preceding lemma, we are going to build a noisy connectivity matrix, whose connection values are of the same order of magnitude of the connectivity matrix, and that we will sum to the deterministic matrix.

The first point to consider is that if we put to zero, at random,  $1 - p$  fraction of the components of a matrix, then the average value of the post synaptic currents will decrease. So, if we strengthen the connections by a factor  $1/p$  and then we randomly set to zero connections with probability  $1-p$ , we will obtain a sparse matrix whose post-synaptic currents have the same expected values as the fully connected one.

$$M^* \leftarrow M/p$$

$$M^* \leftarrow M^* \otimes (\text{rand}([N, N]) < p)$$

$\otimes$  is the point-wise matrix multiplication.

The stochastic matrix is:

$$F = \alpha \cdot \frac{M}{p} \otimes (\text{rand}([N, N]) < p) \cdot (I - u_1 \cdot u'_1 - u_2 \cdot u'_2 - u_3 \cdot u'_3)$$

$\alpha = \frac{R_2}{R_1}$  is a parameter that we will determine in the next section.

#### 3.10.7.2 Rescaling the spectral radius in sparsely connected matrices

For the sake of this argument, we will call  $Q$  the part of the deterministic matrix that is  $O(1)$ :  $M = \frac{1}{\sqrt{N}}Q$ . The spectral radius  $R$  of a matrix  $L_{ij}$  sampled with a probability  $p$  is related to the mean variance of the connection strength through:

$$\frac{R^2}{N} = p(1-p) \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N L_{ij}^2$$

Taking  $L = M \otimes (\text{rand}([N, N]) < p)/p$ , we compute a the associated spectral radius  $R_1$

$$R_1^2 = p(1-p) \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N Q_{ij}^2 \otimes (\text{rand}([N, N]) < p)/p$$

Having sampled at random a matrix with Bernoulli noise might result in a matrix whose spectral radius is higher than one, which makes the system unstable. If one wants to rescale the spectral radius from  $R_1$  to  $R_2$ , one has to multiply the connectivity (both F and M) by  $\frac{R_2}{R_1}$ , so that the new spectral radius (to the square) is:

$$R^2 = p(1-p) \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{R_2}{R_1}\right)^2 Q_{ij}^2 \otimes (\text{rand}([N, N]) < p)/p = R_1^2 \cdot \left(\frac{R_2}{R_1}\right)^2 = R_2^2$$

We multiply also M because we want that the noise on the connections be of the same order as the connections in the deterministic matrix.

### 3.10.7.3 Network size in sparsely connected matrices

One consequence of the formula of the spectral radius :

$$\frac{R^2}{N} = p(1-p) \left[ \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N L_{ij}^2 \right]$$

is that if we expand the dimensionality  $N$  of the matrix  $L$  by "recopying" the elements as we increase the size, the mean value of the connection  $\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N L_{ij}^2$  stays the same, and then the spectral radius also stays the same.

However, as we increase the dimension, the value of the eigenvalues increases -as  $O(\sqrt{N})$ . In the last paragraph we discussed the re-normalisation of the connectivity, in order to have a target spectral radius  $R_2$  that is not greater than the unity.

When we perform such re-normalisation, the eigenvalues decrease by a factor  $\frac{R_2}{R_1}$ .

$$\lambda \leftarrow \frac{R_2}{R_1} \lambda$$

One can then increase the dimensionality to a new dimension  $N$  after having normalized, in order to compensate for the decrease of the eigenvalues.

$$\lambda \leftarrow \frac{R_2}{R_1} \sqrt{N} \lambda$$

We then obtain the following equality:

$$N = \left(\frac{R_1}{R_2}\right)^2$$

$N$  is the suitable network size of our network, so that the network has the eigenvalues we want, and we want them to be the same ones as in low dimensions, and also in such a way that the spectral radius is equal to the spectral radius we want  $R_2 < 1$ .

Another consequence from this equality is that the bigger the target spectral radius  $R_2$ , the smaller the matrix dimension has to be.

#### 3.10.8 Summary of the implementation of the high dimensional model

The purpose of this part is to put together all the different steps needed to implement the high dimensional version of the TCA model.

In order to implement the model, we first choose in low dimension a suitable parameter regime. Then, we make a first draft high dimensional system, of for example 100 neurons, with a deterministic and a stochastic part. We can then compute the spectral radius  $R_1$  of the noise matrix, which is independent of the network size. After choosing a target ratio  $R_2$ , we determine the suitable network size of the network:  $N = \left(\frac{R_1}{R_2}\right)^2$ .

Finally, we make a network  $J$  of dimension  $N$ , with a deterministic and a stochastic part  $J = (M + F)$ , having applied the condition  $F \leftarrow F(I - \sum u_i \cdot u_i')$ . Finally, we multiply by the scaling  $\frac{R_2}{R_1} : J \leftarrow \frac{R_2}{R_1} J$ .

### 3.10.8.1 Algorithm

The recurrent network follows the linear stochastic equation:

$$\dot{X} = -X + JX + \xi$$

Definitions:  $J \in R^{N \times N}$ ,  $\xi.dt|_{i=1,\dots,N} \sim N(0, 1)$   $J$  is the connectivity matrix.

**a) Low dimensional model: Two excitatory populations and one inhibitory population.**

$$T = \begin{pmatrix} \lambda_C & W & 0 \\ 0 & \lambda_U & V \\ 0 & 0 & \lambda_D \end{pmatrix}$$

$$u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad u_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad u_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

$$J = [u_1, u_2, u_3] \cdot T \cdot [u_1, u_2, u_3]^t$$

The vectors  $u_1, u_2, u_3$  are called respectively the competitive, the uniform and the difference.

**b) High dimensional model**

The high dimensional version of the connectivity matrix shows a graded connectivity. We don't have anymore, like in the LD version of the model, two homogeneous groups of excitatory cells. The gradation of the connectivity is given by the competitive vector  $u_1$  in dimension  $N$ . In addition of the graded connectivity, the HD version of the model is not simply a low rank representation in high dimensions of a low dimensional model. In fact our high dimension is full rank, and can be decomposed as a sum of a noisy part  $F$  and of a deterministic part  $M$ :

$$J = \alpha(M + F) \quad \alpha \text{ is a real scalar}$$

The noisy matrix  $F$  is composed by elements drawn from a Bernoulli distribution, built in such a way that if we averaged over realisations of the

noise, the mean value  $\langle F_{(i,j)} \rangle$  of the noisy component of the connectivity from the cell  $j$  to  $i$ , would be 0. We will first show how to build the connectivity matrix as if we knew the dimensionality  $N$ , and then we will see how to determine it. The scalar factor  $\alpha$  is related with the normalisation of the spectral radius, as we will see subsequently.

#### b.1) Deterministic matrix $M$

$$u_1 = \frac{1}{||u_1||} \begin{pmatrix} 1^\kappa \\ 0.9^\kappa \\ \dots \\ 0 \\ 0 \\ \dots \\ -0.9^\kappa \\ -1^\kappa \\ 0 \\ \dots \\ 0 \end{pmatrix} \quad u_2 = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 \\ 1 \\ \dots \\ \dots \\ \dots \\ 1 \\ 1 \end{pmatrix} \quad u_3 = \frac{1}{\sqrt{f + (1-f)L^2}} \begin{pmatrix} 1 \\ \dots \\ 1 \\ 1 \\ \dots \\ 1 \\ -L \\ \dots \\ -L \end{pmatrix}$$

$L = \frac{f}{1-f}$ . The graded competitive high dimensional competitive vector  $u_1$  is defined by: **1:** creating a linear progression from 1 to 0 of  $\frac{f \cdot N}{2}$  elements. **2:** elevating element by element this progression to some power  $\kappa$ . **3:** filling the vector  $u_1$  with the new progression and the inverted progression, and zero padding the vector until attaining  $N$ . **4:** Normalizing the vector.

The proportion of excitatory neurons is  $f$ . The difference vector  $u_3$  is such that  $\sum_i u_3|_i = 0$ .

The connectivity matrix is:

$$M = \sqrt{N} \cdot [u_1, u_2, u_3] \cdot T \cdot [u_1, u_2, u_3]^t$$

With this scaling,  $M \sim O(\frac{1}{\sqrt{N}})$ .

#### b.2) Noisy connectivity matrix $F$

$$F = (M - \frac{1}{p}M \otimes [\text{rand}([N, N]) < p]).(I - \sum_{i=1}^3 u_i \cdot u'_i)$$

$\otimes$  describes a point-wise matrix product.  $p$  is the connection probability and  $[\text{rand}([N, N]) < p]$  is a  $N$  by  $N$  matrix whose elements are one with probability  $p$  and zero with probability  $1 - p$ . The matrix multiplication by the factor  $(I - \sum_{i=1}^3 u_i \cdot u'_i)$  is a condition that we impose so that the eigenspectrum of the sum matrix  $M+F$  is the superposition of the eigenspectrum of  $M$  and  $F$ .

### b.3) Determining the size $N$ of the high dimensional system

The strategy to determine the size of the matrix consists in first creating a mock matrix  $M$  of an arbitrary dimension  $N = 100$ , with the same parameters. The spectral radius of the noise matrix is computed as:

$$R = \sqrt{N \cdot p(1-p) \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (\frac{1}{p} M_{(i,j)})^2}$$

We can re-normalize the connectivity in such a way that the spectral radius attains a target  $R_{target}$ .

The network size is determined as:

$$N = \lfloor (\frac{R}{R_{target}})^2 \rfloor$$

$\lfloor x \rfloor$  is the floored round value of  $x$ . The scalar factor  $\alpha$  of the definition of  $J$  is simply:

$$\alpha = \frac{1}{\sqrt{N}}$$

We then rebuild the network according to this size  $N$ .

### Numerical values:

$\lambda_C = 0.69$ ,  $\lambda_U = -3$ ,  $\lambda_D = 0.6$ ,  $W = 1$ ,  $V = 19$ ,  $f = 0.8$ ,  $p = 0.2$ ,  $R_{target} = 0.79$ ,  $\kappa = 1.05$

### 3.10.9 Translation-invariant connectivity matrices

#### 3.10.9.1 Properties of translation invariant symmetric matrices

Circulant matrices are square  $N \times N$  matrices for which  $W_{ij} = w(j-i)$ , where  $w$  is an  $N$  periodic function  $w(a+N) = w(a)$ . For circulant symmetric matrices  $w(a) = w(-a)$ .

A very remarkable property of every circulant matrix of dimension  $N$  is that its eigenvectors are always the same. We will show that the union of two sets of vectors  $\{p\}$  and  $\{q\}$  constitute the basis of eigenvectors of any circulant matrix. The  $j^{th}$  component of the  $l^{th}$  vector from  $\{\vec{p}\}$  (respectively  $\{\vec{q}\}$ ) is:

$$\vec{p}^l|_j = \cos\left(\frac{2\pi jl}{N}\right), l \in [0, \dots, \text{floor}\left(\frac{N}{2}\right)] \quad \vec{q}^l|_j = \sin\left(\frac{2\pi jl}{N}\right), l \in [1, \dots, \text{floor}\left(\frac{N-1}{2}\right)]$$

Before show this, lets just recall the definition of the discrete Fourier transform of a  $N$  periodic function  $w$  :

$$\begin{aligned} \tilde{w}_{(l)} &= \sum_{r=0}^{N-1} e^{\frac{-2\pi i l r}{N}} w(r) = \sum_{r=0}^{N-1} w(r) \cos\left(\frac{2\pi l r}{N}\right) + i \cdot \sum_{r=0}^{N-1} w(r) \sin\left(\frac{-2\pi l r}{N}\right) \\ &= \text{Real}(\tilde{w}_{(l)}) + i \text{Imag}(\tilde{w}_{(l)}) \end{aligned}$$

Now, lets calculate for  $\vec{p}^{(l)}$  and  $\vec{q}^{(l)}$ :

$$\begin{aligned} (W \cdot \vec{p}^{(l)})|_k &= \sum_j W_{kj} \cos\left(\frac{2\pi jl}{N}\right) \\ &= \sum_j w_{(j-k)} \cos\left(\frac{2\pi jl}{N}\right) \\ &= \sum_r w_{(r)} \cos\left(\frac{2\pi(r+k)l}{N}\right) \\ &= \sum_r w_{(r)} \cos\left(\frac{2\pi r l}{N}\right) \cos\left(\frac{2\pi k l}{N}\right) - \sum_r w_{(r)} \sin\left(\frac{2\pi r l}{N}\right) \sin\left(\frac{2\pi k l}{N}\right) \\ &= \vec{p}^{(l)}|_k \cdot \text{Real}(\tilde{w}_{(l)}) + \vec{q}^{(l)}|_k \cdot \text{Imag}(\tilde{w}_{(l)}) \end{aligned}$$

$$(W.q^{(l)})|_k = -Imag(\tilde{w}_{(l)}) \cdot p^{(l)}|_k + Real(\tilde{w}_{(l)}) \cdot q^{(l)}|_k$$

$$\begin{cases} W.\vec{p}^l = Real(\tilde{w}_{(l)}) \cdot \vec{p}^l + Imag(\tilde{w}_{(l)}) \cdot \vec{q}^l \\ W.\vec{q}^l = -Imag(\tilde{w}_{(l)}) \cdot \vec{p}^l + Real(\tilde{w}_{(l)}) \cdot \vec{q}^l \end{cases}$$

For any circulant matrix, the complex eigenvectors are  $\xi^l = \vec{p}^l \pm i \cdot \vec{q}^l$ . The eigenvalues of a circulant matrix  $W$  depend on the discrete Fourier transform of the characteristic periodic function  $w$ .

$$\lambda_{(l)} = Real(\tilde{w}_{(l)}) \pm i \cdot Imag(\tilde{w}_{(l)})$$

In the case of circulant symmetric matrices, these expressions are simplified. As every symmetric matrix has real eigenvalues and real eigenvectors, if  $W$  is circulant symmetric, it is easy to show that  $Imag(\tilde{w}_{(l)}) = 0$  then

$$\begin{cases} W.\vec{p}^l = Real(\tilde{w}_{(l)}) \cdot \vec{p}^l \\ W.\vec{q}^l = Real(\tilde{w}_{(l)}) \cdot \vec{q}^l \end{cases}$$

Note however that  $\vec{q}^0 = \vec{0}$  and if  $N$  is even we also have  $\vec{q}^{N/2} = \vec{0}$ . So if  $N$  is even, there are  $\frac{N}{2} + 1$  eigenvectors from the  $\{p\}$  basis ( $l = 0, \dots, \frac{N}{2}$ ) and  $\frac{N}{2} - 1$  from the  $\{q\}$  basis ( $l = 1, \dots, \frac{N}{2} - 1$ ). The first eigenvector is always the uniform  $\vec{p}^0 = \frac{1}{\sqrt{N}} \vec{1}$ , with corresponding eigenvalue,  $\tilde{w}_{(0)} = \sum_n w_{(n)}$ .

Still, if  $N$  is even, the last eigenvector  $\vec{p}^{N/2} = \tilde{w}_{(l)}(1, (-1), 1, \dots, (-1))'$  and the corresponding eigenvalue is  $\sum_n w_{(n)}(-1)^n$ . All the other eigenvectors  $\vec{p}^l$  and  $\vec{q}^l$  share the same eigenvalue:  $Real(\tilde{w}_{(l)})$ .

$$\begin{cases} \text{eigenvectors : } \{p^0, p^1, q^1, p^2, q^2, \dots, p^{N/2}\} = \{e_1, e_2, e_3, e_4, \dots, e_N\} \\ \text{eigenvalues : } \{\lambda_0, \lambda_1, \lambda_1, \lambda_2, \lambda_2, \dots, \lambda_{\frac{N}{2}}\} = \{\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3, \dots, \tilde{\lambda}_N\} \end{cases}$$

### 3.10.9.2 Translation invariant connectivity matrices lead to independent two by two matrices for each spatial frequency

We just saw that all translation invariant matrices have the same basis of eigenvectors, independently of the spatial frequential content of each matrix. These eigenvectors are orthogonal, because each translation invariant matrix



is symmetric. Following Murphy and Miller's argument, let's consider the Fourier basis :

$$B = \{e_1^E, e_1^I, \dots, e_N^E, e_N^I\} \quad e_1^E = \begin{pmatrix} e_1 \\ 0 \end{pmatrix} \quad e_1^I = \begin{pmatrix} 0 \\ e_1 \end{pmatrix}$$

$\{e_i\}_{i=1..N}$  is the eigenvector basis of a N dimensional circulant matrix. The basis B is orthonormal, but is not the eigenbasis of J. We call  $P_B$  the associated (normalized) rotation matrix. J can be expressed in the basis B as:

$$F = P'_B \cdot J \cdot P_B = \begin{pmatrix} \tilde{\lambda}_1^{EE} & \tilde{\lambda}_1^{EI} & 0 & 0 & \dots & \dots & 0 & 0 \\ \tilde{\lambda}_1^{IE} & \tilde{\lambda}_1^{II} & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & \tilde{\lambda}_2^{EE} & \tilde{\lambda}_2^{EI} & \dots & \dots & 0 & 0 \\ 0 & 0 & \tilde{\lambda}_2^{IE} & \tilde{\lambda}_2^{II} & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \dots & \tilde{\lambda}_N^{EE} & \tilde{\lambda}_N^{EI} \\ 0 & 0 & 0 & 0 & \dots & \dots & \tilde{\lambda}_N^{IE} & \tilde{\lambda}_N^{II} \end{pmatrix}$$

We will make the choice in all this section that the profile  $w_{XY}$  is the same for all four quadrants of J- up to a multiplicative constant:  $w_{EI}(l) = -\alpha_1 \cdot w_{EE}(l)$ ,  $w_{IE}(l) = \alpha_2 \cdot w_{EE}(l)$  and  $w_{II}(l) = -\alpha_3 \cdot w_{EE}(l)$ ,  $\alpha_{1,2,3} > 0$ . The frequential content is then also the same at all frequencies, up to those proportionality constants. An additional constraint that we will impose is that the uniform vector is an eigenvector of J, the connectivity matrix. The consequence of this is that for each 2x2 sub-matrix of F, the uniform vector of dimension two will also be an eigenvector, so that:  $\alpha_3 = \alpha_2 + \alpha_1 - 1$ . This will simplify considerably the equations (see Schur decomposition of a 2 by 2 matrix) and will allow us to have an analytical grip on the variance of the modes as a function of the frequency. Given these choices in the Fourier base, we have sub-matrices which have the uniform as eigenvector, and which are proportional the ones to the others.

$$\kappa(n) \cdot \begin{pmatrix} \tilde{\lambda}_1^{EE} & \tilde{\lambda}_1^{EI} \\ \tilde{\lambda}_1^{IE} & \tilde{\lambda}_1^{II} \end{pmatrix} = \begin{pmatrix} \tilde{\lambda}_n^{EE} & \tilde{\lambda}_n^{EI} \\ \tilde{\lambda}_n^{IE} & \tilde{\lambda}_n^{II} \end{pmatrix}, \quad \tilde{\lambda}_n^{EE} + \tilde{\lambda}_n^{EI} = \tilde{\lambda}_n^{IE} + \tilde{\lambda}_n^{II}$$

Murphy and Miller advance that if the connectivity strength  $w(n)$  decreases smoothly enough with the distance, then the Fourier coefficients

$\tilde{\lambda}_n^{EE} = \kappa(n)\tilde{\lambda}_1^{EE}$  are likely to be positive ( $\kappa(n) \geq 1, \forall n$ ), and then for each sub-block the first column of is positive and the second column is negative so that the sub-matrices of  $F$  are Dale. This is not always the case. As a counter example, take the function  $f(x) = (1 + x^2)e^{-x^2}$ . This function is even and decreases smoothly with the distance, and it is strictly positive. Its Fourier transform is  $F(\omega) = \frac{(6-\omega^2)e^{-\omega^2/4}}{4\sqrt{2}}$ , so it is negative in  $] \sqrt{6}, +\infty[$ . Notwithstanding, we tried many profiles, like the cosine -which has only one harmonic- or the Gaussian -whose Fourier transform is positive-, and some others to see which ones respect the Dale's condition. Generally, it is hard to find slowly decreasing profiles such that all the Fourier coefficients are positive. Conversely if we set to zero the negative coefficients of those profiles in the Fourier basis and look at the connectivity in the canonical basis, the matrix are not usually strictly Dale, but only approximately.

### 3.10.9.3 From the Fourier basis B to the Schur basis S

In the appendix "Schur decomposition of a 2 by 2 matrix", we can see that a given matrix  $\begin{pmatrix} \tilde{\lambda}_{EE} & \tilde{\lambda}_{EI} \\ \tilde{\lambda}_{IE} & \tilde{\lambda}_{II} \end{pmatrix}$  has two eigenvectors that can be written :  $\begin{pmatrix} 1 \\ x_+ \end{pmatrix}$  and  $\begin{pmatrix} 1 \\ x_- \end{pmatrix}$ . We previously set the constraint that for each 2x2 submatrix, the uniform vector is an eigenvector of these sub-matrices:  $(\tilde{\lambda}_{EE} + \tilde{\lambda}_{EI} = \tilde{\lambda}_{IE} + \tilde{\lambda}_{II})$ .

Given this constraint, we have simply:

$$x_+ = 1 \quad \text{and} \quad x_- = \frac{\tilde{\lambda}^{IE}}{|\tilde{\lambda}^{EI}|}$$

With regard to the basis B, these eigenvectors  $Eig_B$  can be written:

$$Eig_B = \left\{ \left\{ \begin{pmatrix} 1 \\ x_+^1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ x_-^1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{pmatrix} \right\}, \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ x_+^2 \\ 0 \\ 0 \\ \dots \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ x_-^2 \\ 0 \\ 0 \\ \dots \end{pmatrix} \right\}, \dots \right\}$$

These eigenvectors are not orthogonal. When we orthogonalise them in the basis B, we are going to obtain the new Schur basis S, with its associated rotation matrix  $P_S$ :

$$S_B = \left\{ \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{pmatrix} \right\}, \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ \dots \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ \dots \end{pmatrix} \right\}, \dots \right\}$$

Each pair of eigenvectors can be indexed with the number of the block in the Fourier basis

$$\begin{cases} \lambda_+^n &= \tilde{\lambda}_n^{EE} + \tilde{\lambda}_n^{EI} \\ \lambda_-^n &= \tilde{\lambda}_n^{EE} - \tilde{\lambda}_n^{IE} \\ t^n &= \tilde{\lambda}_n^{IE} - \tilde{\lambda}_n^{EI} \end{cases}$$

In the Schur basis, the connectivity matrix is then:

$$T = P'_S \cdot F \cdot P_S = P'_S \cdot P'_B \cdot J \cdot P_B \cdot P_S = \begin{pmatrix} \tilde{\lambda}_+^1 & t^1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & \tilde{\lambda}_-^1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & \tilde{\lambda}_+^2 & t^2 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & \tilde{\lambda}_-^2 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \dots & \tilde{\lambda}_+^N & t^N \\ 0 & 0 & 0 & 0 & \dots & \dots & 0 & \tilde{\lambda}_-^N \end{pmatrix}$$

However, given as we saw in the first part that eigenvectors of translation invariant matrices come in pairs of cosines and sines, with the same eigenvalue (excepting the first eigenvector), we have  $\tilde{\lambda}_2^{XY} = \tilde{\lambda}_3^{XY}$  and  $\tilde{\lambda}_4^{XY} = \tilde{\lambda}_5^{XY}$  and so on. Then, we also have that the numerical values (the eigenvalues and the feed-forward link) of the second Schur decomposition are equal to the values of the third one, the fourth to the fifth and so on.

In the canonical basis, the Schur vectors can be written as:

$$S_c = \{ \{ (e_1^E + e_1^I), (e_1^E - e_1^I) \}, \{ (e_2^E + e_1^I), (e_2^E - e_2^I) \}, \dots \}$$

$$S_c = \{ \left\{ \begin{pmatrix} 1 \\ \dots \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ \dots \\ 1 \\ -1 \end{pmatrix} \right\}, \left\{ \begin{pmatrix} e_2 \\ e_2 \end{pmatrix}, \begin{pmatrix} e_2 \\ -e_2 \end{pmatrix} \right\}, \dots \}$$

We then see why the first feed forward pair corresponds to the difference to uniform, and the other pairs have Schur vectors called minus and plus.

### 3.10.9.4 Variance of the Schur modes as a function of the frequency

Starting from the canonical basis and going directly to the Schur basis, differential equation characterizing the system might be written:

$$\dot{X} = -X + JX + \xi$$

$$\dot{Y} = -Y + TY + \xi \quad T = P'_S \cdot P'_B \cdot J \cdot P_B \cdot P_S$$

When we compute the Lyapunov equation to obtain the covariance matrix of the Schur decomposition, we can make it directly at the level of the blocks, because T is block diagonal. Given a connectivity matrix  $T_n = \begin{pmatrix} \lambda_+^n & t^n \\ 0 & \lambda_-^n \end{pmatrix}$ , solving the Lyapunov equation:

$$(T_n - I_2)C_n + C_n(T_n - I_2)' = -\sigma^2 \cdot I_2$$

gives the variances v and covariances c matrix between the modes:

$$\begin{cases} v_+^n &= \frac{-\sigma^2}{2(\lambda_+^n - 1)} \left( 1 + \frac{(t^n)^2}{(\lambda_-^n - 1)(\lambda_-^n + \lambda_+^n - 2)} \right) \\ v_-^n &= \frac{-\sigma^2}{2(\lambda_-^n - 1)} \\ c_{+-}^n &= \frac{t^n \cdot \sigma^2}{2(\lambda_-^n - 1)(\lambda_-^n + \lambda_+^n - 2)} \end{cases}$$

Given the imposed constraints,  $\lambda_{\pm}^n = \kappa(n)\lambda_{\pm}^1$  and  $t^n = \kappa(n)t^1$ .

We then have the formula:

$$\begin{cases} v_+^n &= \frac{-\sigma^2}{2(\kappa(n)\lambda_+^1-1)} \left(1 + \frac{(\kappa(n)t^1)^2}{(\kappa(n)\lambda_-^1-1)(\kappa(n)\lambda_-^1+\kappa(n)\lambda_+^1-2)}\right) \\ v_-^n &= \frac{-\sigma^2}{2(\kappa(n)\lambda_-^1-1)} \\ c_{+-}^n &= \frac{\kappa(n)t^1 \cdot \sigma^2}{2(\kappa(n)\lambda_-^1-1)(\kappa(n)\lambda_-^1+\kappa(n)\lambda_+^1-2)} \end{cases}$$

Replacing the values of  $\lambda_{\pm}^1$  and  $t^1$ :

$$v_+^n = \frac{-\sigma^2}{2(\kappa(n)(\tilde{\lambda}_1^{EE} - |\tilde{\lambda}_1^{EI}|) - 1)} \left(1 + \frac{(\kappa(n)(\tilde{\lambda}_1^{IE} + |\tilde{\lambda}_1^{EI}|))^2}{(\kappa(n)(\tilde{\lambda}_1^{EE} - \tilde{\lambda}_1^{IE}) - 1)(\kappa(n)(2\tilde{\lambda}_1^{EE} - \tilde{\lambda}_1^{IE} - |\tilde{\lambda}_1^{EI}|) - 2)}\right) \quad (1)$$

Therefore, we can examine how does the variance of the plus mode evolves with the frequency. As  $n \rightarrow N$ ,  $\kappa(n) \rightarrow 0$ , because the power of a smoothly decreasing profile has low power in the high frequencies. The high frequency components of the system (which are uncoupled from the other frequencies), behave like a subsystem that is uncoupled, having no recurrent connections:  $\dot{Y} = -Y + \xi$ . Consequently, the variance of the last modes is given approximately by the variance of the noise:  $v_+^n = \frac{\sigma^2}{2}$ .

We can imagine two situations:

- The first spatial frequencies that compose the recurrent connectivity strongly quench the dynamics, so that the variance associated with the first modes is even smaller than the variance of the noise. When we examine the variance of the plus mode as a function of the frequency, we expect the variance to increase.

- The first frequencies in the connectivity have an amplifying effect so that the variance associated to the plus mode is higher than the noise variance. Coarsely, we expect the variance of the noise to decrease as a function of the frequency. The question is whether we can find a parameter regime in which there is a transient increase in the variance, to then decay towards  $\frac{\sigma^2}{2}$ .

The sum mode is composed by two summands:

$$\frac{\sigma^2}{-2(\lambda_+^1 \kappa(m) - 1))} \quad \text{and} \quad \frac{\sigma^2}{-2((\lambda_+^1 \kappa(m) - 1))} \frac{(t^1 \cdot \kappa(m))^2}{((\lambda_-^1) \kappa(m) - 1)((\lambda_+^1 + \lambda_-^1) \kappa(m) - 2)}$$

The first term decreases (or increases) monotonically with the frequency.

The second term can increase in two situations: first, if the denominator grows faster than the denominator and second if the denominator decreases faster than the numerator. The first alternative is not plausible, because  $\kappa(n)$  decreases as a function of the frequency- . For the second alternative, it is not excluded that it might happen because on the denominator we have a damping term to the cube  $\kappa(n)^3$  whereas in the numerator we have a damping term to the square  $\kappa(n)^2$ .

What could be the mechanism behind such kind of amplification ? Lets imagine that we have a first harmonic whose sub-block matrix is proportional to the offset. In the Schur domain, taking into account the decay, the sub-block equations are linked through:

$$\kappa(2) \cdot \begin{pmatrix} \tilde{\lambda}_+^1 & t^1 \\ 0 & \tilde{\lambda}_-^1 \end{pmatrix} - I_2 = \begin{pmatrix} \tilde{\lambda}_+^2 & t^2 \\ 0 & \tilde{\lambda}_-^2 \end{pmatrix} - I_2$$

Recalling that the eigenvalues are negative, multiplying a negative quantity by a positive number smaller than one results in a bigger quantity.

$$0 > \kappa(2) \tilde{\lambda}_+^1 > \tilde{\lambda}_+^1$$

If the amount of negative feed-back is a bit smaller in the first harmonic (of index 2) with respect to the amount of negative feed-back of the dc (index 1):  $\tilde{\lambda}_+^1 - 1 > \kappa(2) \tilde{\lambda}_+^1 - 1$ , then by normal amplification it could be that the first harmonic displays more variance that the DC. However, it doesn't needs to happen always, because as the damping  $\kappa(n)$  decreases the feed-forward coupling  $t^1 \kappa(n)$  decreases and the patterns plus and minus become uncoupled.

### 3.10.9.5 Examples of competitive activity with translation invariant matrices

In order to illustrate that indeed it is possible to have competition, with a Gaussian profile. As we can see in figure 3.17 E, we see a transient increase

in the variance of the modes  $c_+$ , that then decreases to 1, half of the noise variance  $\frac{\sigma^2}{2}$ .

The matrix parametrisation is the following:

$$\begin{pmatrix} J_{EE} & J_{EI} \\ J_{IE} & J_{II} \end{pmatrix} = \begin{pmatrix} J_{EE} & -\gamma_1 J_{EE} \\ \gamma_2 J_{EE} & -(\gamma_1 + \gamma_2 - 1) J_{EE} \end{pmatrix}, \quad J_{EE}|_{(i,j)} = K e^{\frac{-d_{(i,j)}^2}{w^2}}$$

$$d_{(i,j)} = \begin{cases} j - i & \text{if } i - j < \frac{N}{2}; \\ N - j + i & \text{else.} \end{cases}$$

In figure 3.17 E, we plotted the formula (1) and also the measured variance.

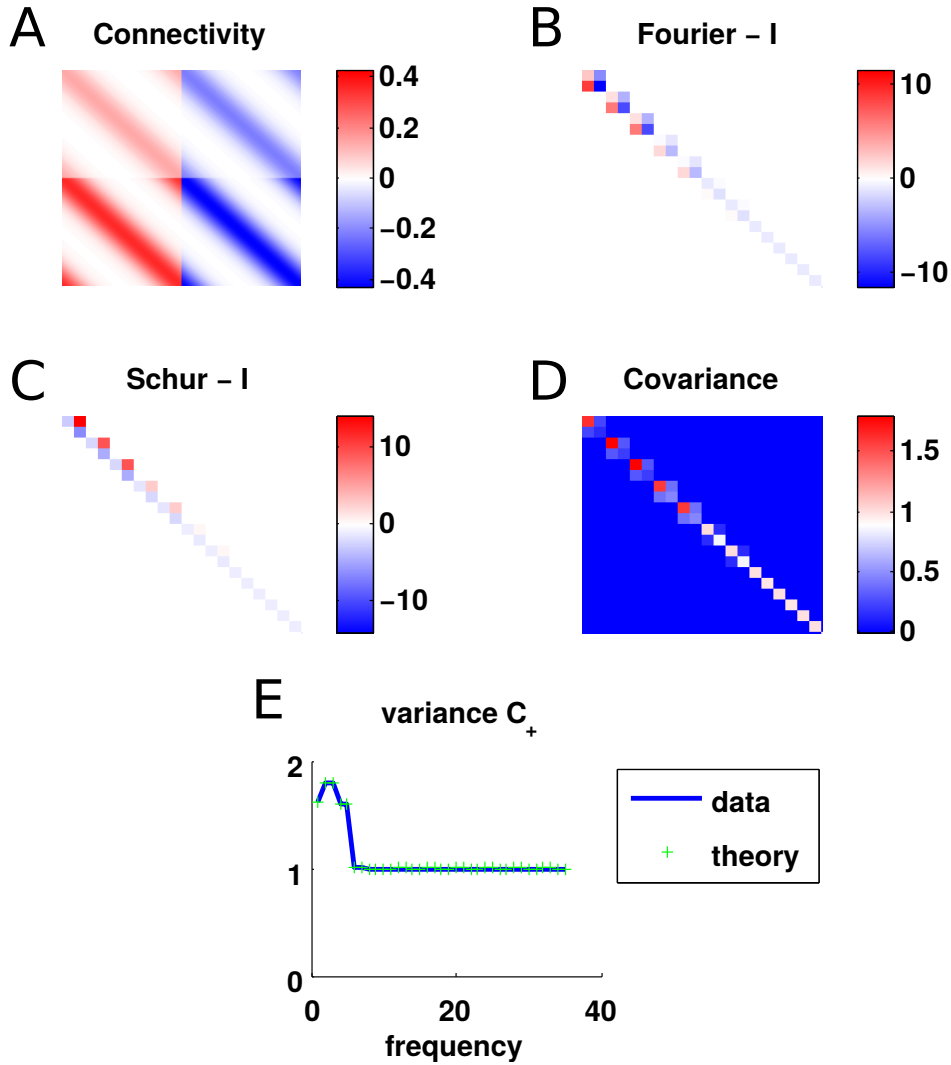


Figure 3.17: Example of a connectivity matrix with a gaussian profile that generates more variance on the first plus mode than in the uniform. A: connectivity matrix, B: connectivity matrix in the Fourier basis (minus the identity), C: connectivity matrix in the Schur basis (minus the identity), D: covariance in the Schur basis, E: variance of the plus mode as a function of the frequency  $\gamma_1 = 1.5$ ,  $\gamma_2 = 2.5$ ,  $K = 20$ ,  $2N = 140$ ,  $w = \lfloor \frac{N}{5} \rfloor$ ;  $\sigma^2 = 2$



# Chapter 4

## Cortical neurons integrate common inputs from sensory thalamus

### HIGHLIGHTS

- This chapter introduces a technique developed by N. Morgenstern and L. Petreanu, showing how through opto-genetic stimulation and electro-physiological recordings, it is possible to deduce an anatomical characteristic: the spatial selectivity of thalamo-cortical projections in the cortical layers.
- We built a model of the experiment, relating the proportion of shared axons with the way that EPSCs from pairs of neurons covariate.
- This study shows that pairs of connected neurons in L4 tend to receive more shared input from dLGN, than pairs of non connected neurons. The same happens for pairs of connected cells, where one of the cells is in L4 and the other in L2/3.

### Contents

<b>4.1</b>	<b>Introduction: the visual pathway, from the retina to the primary cortex . . . . .</b>	<b>179</b>
------------	--	------------

4.1.1	The Hubel and Wiesel model . . . . .	179
4.1.2	Amplification of the thalamo-cortical projection .	182
4.1.3	Amplification and connectivity . . . . .	183
4.1.4	Aim of this chapter . . . . .	184
<b>4.2</b>	<b>Experiments . . . . .</b>	<b>185</b>
4.2.1	Preliminary experiments: determining the distribution and strength of dLGN axonal projections across cortical layers of V1 . . . . .	185
4.2.2	Probing connectivity between two neurons, measuring the covariations induced by long range projections . . . . .	186
<b>4.3</b>	<b>Modelling the covariance of the membrane potential of a couple of patched cells as a function of the proportion of shared inputs . . . . .</b>	<b>188</b>
4.3.1	Modelling the correlation as a function of the stimulation power . . . . .	188
4.3.2	Modelling the mean correlation . . . . .	191
4.3.3	Differences in mean correlation reflect differences in the proportion of shared axons . . . . .	193
4.3.4	Adding variability about the number of recorded sites in one location . . . . .	197
4.3.5	Examining the consequences of variability in the post synaptic current amplitude . . . . .	198
<b>4.4</b>	<b>Stimulating at only one power and counting the fraction of correlated locations . . . . .</b>	<b>201</b>
4.4.1	Numerical simulation reproducing stimulation at constant power over a grid of points . . . . .	201
4.4.2	Experimental results . . . . .	203
<b>4.5</b>	<b>Discussion . . . . .</b>	<b>206</b>
4.5.1	Summary . . . . .	206
4.5.2	Modelling correlations . . . . .	206
4.5.3	Connectivity and function . . . . .	207
4.5.4	Importance of long range projections to understand cortical functions . . . . .	209

4.6	Appendix	209
4.6.1	Mathematical proofs	209
4.6.2	Numerical simulation	211

---

## 4.1 Introduction: the visual pathway, from the retina to the primary cortex

### 4.1.1 The Hubel and Wiesel model

The retina is a thin layer of tissue situated in the back of the eye. It is covered by two kinds of photo-receptors called rods, in the fovea, and cones in the periphery. After passing through bipolar cells, horizontal cells, amacrine cells, the visual information attains the ganglion cells (see the figure 4.1 B). We now know, thanks to the seminal work of Hubel and Wiesel [72], [73], that the retina detects local contrast of the images, and the brain’s work is to use this information to segment and identify the objects in presence. This strong insight came from the fact that Hubel and Wiesel observed that individual cells in primary visual cortex of cats responded to bars with specific orientations [72]. Subsequently, they proposed a model, a possible implementation of how this orientation specificity could arise, that we will revisit in here.

The visual environment is projected through the lens into the retina, on which an image is formed. In the retina the light is transduced into an electrical signal by the rods, which are sensitive to a specific wavelength range. The rods project onto the bipolar cells. The retina is divided into functional regions called center and surround regions, which correspond to particular regions of the visual space. In center regions of the retina, the bipolar cells project directly to the ganglion cells, exciting them. If on the contrary, the bipolar cells are on a surround region, they project to horizontal cells, which in turn, inhibit the ganglion cells. Overall, if a punctual visual stimulus is physically on a center, it activates the ganglion cells, but if it is on the periphery, it inhibits this ganglion cell. Such a ganglion cell is called an **on-off ganglion cells**, because of its receptive field, and it is represented by a black disc, in which there is a white disk in the middle. Also, there is a dual version of the on-off neurons, called **off-on ganglion cells**, in which

when the stimulus is on the surround, it excites the ganglion cell and inhibits it when it is on the center.

The axons of the ganglion cells go through the optic chiasm projecting to the opposite hemisphere and attain the thalamus, and more precisely the visual thalamus, called the lateral geniculate nucleus (**LGN**), see figure 4.1 A. The LGN is thought to act as a relay between the retina and the primary visual cortex V1, which means that its cells convey the same information as the ganglion cells, so we will call also these cells on-off or off-on cells.

As we said before, cells in V1 have the property of being orientation selective, which means that they fire when a bar of a particular orientation is presented in a particular region of the retinal space. Hubel and Wiesel postulated that the mechanism by which neurons in V1 acquire the property of being orientation selective, is similar to the one that generate on-off cells, i.e. that there is a particular spatial selectivity of the pool of neurons that converge to one cell. More precisely, in the case of V1 cells, the idea is that a cell in V1 only receives inputs from a structure in LGN composed of three columns of on-off and off-on LGN cells (see figure 4.1 C). The first and the third columns are composed only of adjacent off-on cells, and the middle column is composed of on-off cells. The orientation of this structure of on-off, off-on cells in the visual space will determine the orientation selectivity of cells in V1, which we depicted with an oriented coloured bar.

Considering the schematic of the canonical cortical microcircuit in the introductory chapter (figure 1.1), we see that the thalamus projects to almost all the cortical layers, except to L1. All over the depth of the cortical column neurons show similar orientation selectivity. The orientation selectivity varies smoothly across the surface of certain species like in cat visual cortex, but in a totally discontinuous way in rodents - even if the fact of not being adjacent doesn't prevent them to be connected.

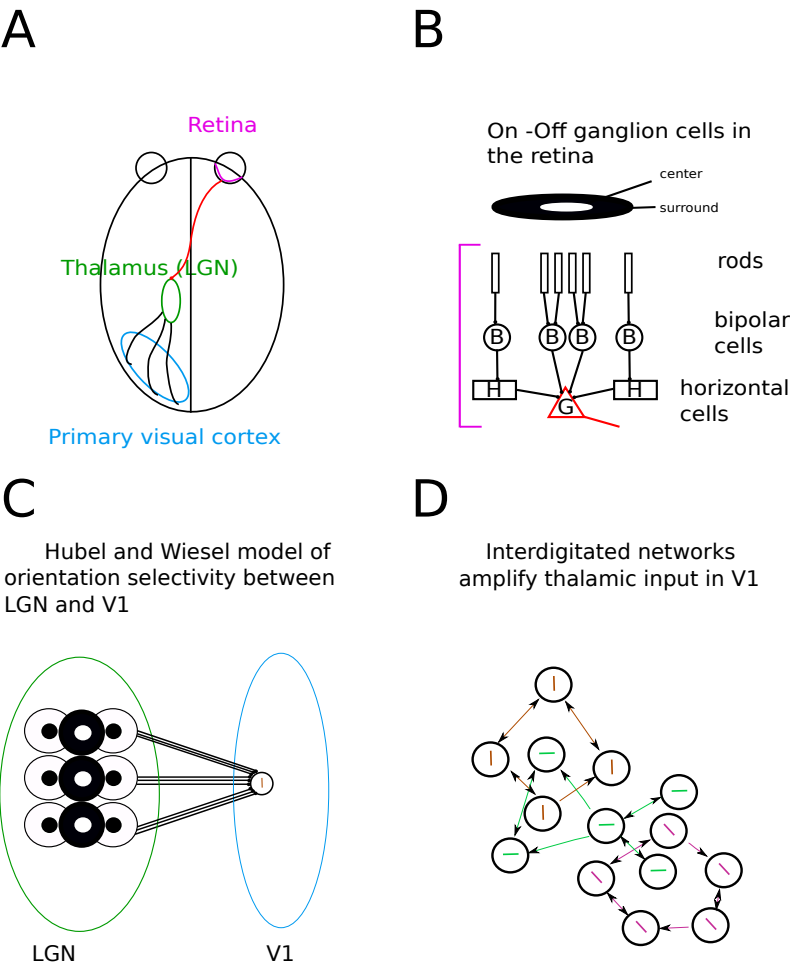


Figure 4.1: A: schematic of the early visual pathway. B: retina stratification. C: Hubel and Wiesel model of orientation selectivity. D: interdigitated cortical subnetworks of neurons with similar stimulus preference.

### 4.1.2 Amplification of the thalamo-cortical projection

Lien and Scanziani [87] used in 2013 whole cell voltage patch to determine how does the tuning of the membrane currents to oriented gratings changed when they silenced layer 4 of mouse V1 with optogenetics.

Briefly, lets recall in what **optogenetics** consists. Inserting a gene from a microalgae in the genome of adeno-associated virus AAV virus, makes the host cell synthesize light-sensitive ion channels. These special ion-channels are called channel-rhodopsins: they open in presence of light, allowing ionic currents to enter into the cell, depolarizing the membrane, and eventually making the neuron spike.

In this experiment, the authors used genetically modified mouse expressing channel-rhodopsin in the parvalbumin-expressing (PV) GABAergic interneurons. The effect of GABA is inhibitory: driving interneurons to spike, inhibits the surrounding excitatory cells. This manipulation isolates the contribution from the thalamus into the cortex, and by default, the contribution of the recurrent connectivity to the tuning of a single cell in L4 (see figure 4.2). Lien and Scanziani discovered that the sub-threshold currents had the same tuning to gratings when the cortex was silenced, as well as when it was not. When light was on, the amplitude of the currents was reduced by about one third.

The conclusion of this experiment, is that the recurrent connectivity is just multiplying and maintaining the thalamic input, which corresponds to the definition of amplification (see scalar case of normal amplification in chapter 3.). The cortical amplification of thalamic input had been already predicted by Douglas et al [38] back in 1995, when they showed that the relative number of synapses reaching a cell cortical cell, was small in comparison with the number of recurrent synapses coming from the cortex: "[In V1] connections arising from the LGN make up less than 10% of the excitatory synapses formed with the spiny stellate cells". They concluded that the signal coming from the thalamus should be amplified somehow, and proposed a computational mechanism to achieve this, namely normal amplification. Normal amplification, as we extensively explained it in the third chapter, appears as we decrease the negative feedback that a single cell or a population of cells exerts onto itself. As Murphy and Miller showed [105], normal amplification is not the only computational mechanism possible: non-normal amplification is very general and might also play a role in

the amplification of joint patterns of correlated activity.

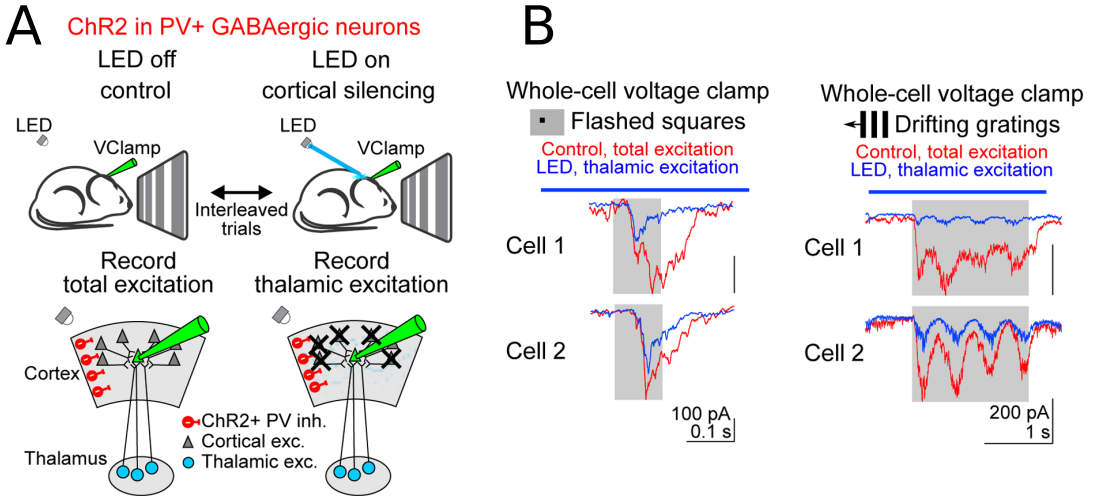


Figure 4.2: Experiment consisting in exciting optically the inhibitory parvalbumin PV+ interneurons in order to silence the excitatory cells, and disentangling the relative contribution from the thalamus and from the cortex to the stimulus-tuning of one cell in L4. A: experimental setup. B: results, subthreshold membrane potential in presence and in absence of optogenetic stimulation, for two different visual stimuli. Figure adapted from [87].

### 4.1.3 Amplification and connectivity

As it is apparent in previous chapters, when we think of a recurrently connected network, as L4 or L2/3, we should not consider that because two excitatory cells are connected, they are going to be positively correlated, because all the other interactions in the circuit also have an impact on their correlation. For example, one cell connected to another can also be strongly connected to an inhibitory cell, which in turn inhibits the second cell. As a rule of thumb, we can nevertheless consider that if the strength of the connection between two cells is very high, and the average strength of synapses is low, these cells are probably going to be correlated. Cossell and colleagues [29] recently showed, using in vivo imaging experiments and then whole-cell recordings in vitro, that cells in layer 2/3 with similar receptive fields tended

to be strongly connected in sub-networks, and that the most correlated cells tended to be bidirectionally connected. Intuitively it is patent that bidirectional strong connections between a couple of cells creates positive feed-back between them, so that when these neurons are perturbed, they are active more time than cells in isolation. We can see examples of this this maintained activity in figure 4.2 B.

We shouldn't consider neither that because two cells are unconnected, they are not going to be correlated, because they can receive common input. Nevertheless, being connected and receiving common input, is going to make them be more positively correlated. Yoshimura et al [164], have shown that pairs of cells that are connected in layer 2/3 tend to receive axons projecting to both cells coming from L4. We say that these pairs of cells receive **shared inputs**.

### 4.1.4 Aim of this chapter

On one side, the Hubel and Wiesel model makes a prediction of the spatial selectivity of convergence of thalamic projections into L4 cortical neurons. On another side, there is evidence of amplification of the thalamo-cortical signal via recurrent connections, in which cells with similar receptive field are strongly connected. Consequently, this work tests the prediction that pairs of cells that are connected in a single cortical layer tend to receive more shared input from the dorsal lateral geniculate nucleus dLGN. Given that the projections from dLGN target many layers, we will also consider whether single axons tend to innervate connected pairs of cells in different layers.

This study was done in collaboration with Nicolás Morgenstern and Leopoldo Petreanu, which conceived the study. The experiments and the data analysis were done by N. Morgenstern. My contribution onto this study is limited to the modelling part. Some of the results presented here were published as a research article [104].

We will first explain the experiments done by N. Morgenstern. Then, we will show two alternative ways in which we modelled this experiment. After that, we will discuss which way was retained for the publication to observe the data, and finally we will see the data and comment the results.



## 4.2 Experiments

### 4.2.1 Preliminary experiments: determining the distribution and strength of dLGN axonal projections across cortical layers of V1

#### 4.2.1.1 Density of thalamocortical projections to different cortical layers

In order to observe the projections from the thalamus to the cortex, one can inject a virus called "AAV-2/1-CAG-Channelrhodopsin-2-Venus", which inserts a gene in the host cells. This gene express a protein called **GFP**, green fluorescent protein which lies in the cytoplasm. When exposed to blue light, this protein called Venus emits back a green light at around 550 nm, allowing to see the somata, the dendrites and the axons.

In all the described experiments, mice of about 15 days old were infected with a virus in the dLGN. Four to twelve days after the infection, mice were euthanized, and acute coronal slices of the primary visual cortex were cut using a vibratome.

The slice were illuminated with a laser. Using a confocal microscope, we can take a picture of the fluorescence of the projecting axons. When the fluorescence is quantified, it appears that, as reported in the literature [59], the densest projections are seen in layer 4, followed by layer 2/3 and some sparser projections to L5 and L6.

#### 4.2.1.2 Strength of thalamo-cortical projections to different cortical layers

The previous virus also leads to the expression of channel-rhodopsin ion channels in the cell axonal membranes and over the axon terminals. In order to measure the connectivity strength between the dLGN projecting axons and cells from different cortical layers, tetrodotoxin (TTX) was applied in order to avoid the recruitment of voltage gated sodium  $Na^+$  channels and to evoke only mono-synaptic responses.

Previous work by Petreanu [115] showed that these axons can be stimulated by light driving, even in the case in which the axons were cut from their soma. When light was applied, axon terminals released the synaptic vesi-

cles. Using whole cell recordings on a post-synaptic cell, the total amount of EPSC's were recorded and summed by for each neuron (for example over the dendritic trees of L2/3), which allowed to quantify the connections strengths and to compare them with respect with the connection strengths of an L4 cell recorded in the same slice. This method is called sub-cellular ChR2-assisted circuit mapping, sCRACM [115].

This experiment showed that dLGN inputs innervate mainly L4 and also lower L2/3 cells, with an average strength -in the first 100  $\mu m$ - slightly smaller than L4 (85%). The thalamic inputs into L5 where only 16% as strong as those contacting L4.

Consequently, on the next experiment, the thalamo-cortical projections that where studied where the ones targeting L4 and L2/3. In the next paragraph we will describe the main experiment. This technique, developed by N. Morgenstern and L. Petreanu, is innovative it is applicable to measure whether long-range axons targets preferentially connected pairs of cells.

### 4.2.2 Probing connectivity between two neurons, measuring the covariations induced by long range projections

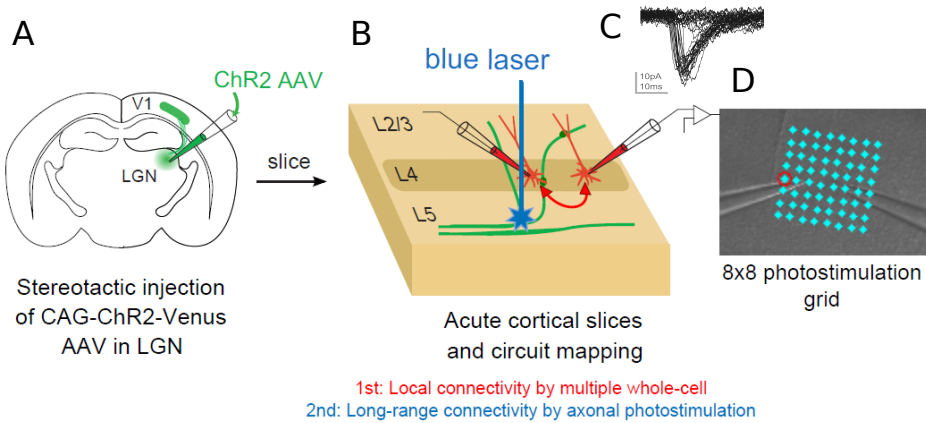
The first step of the procedure is to patch simultaneously, using two electrodes, a pair of cells (see figure 4.3 B). Once patched, it is possible to inject a step of current in one of the two cells to make it spike and observe if this elicits **eEPSCs** (excitatory postsynaptic currents) in the other neuron (figure 4.3 C, see also figure 1.2). If it is the case, we consider that these neurons are connected unidirectionally. We then repeat the same procedure inverting the role of each electrode to probe the existence of the reciprocal connection.

After having determined whether the two cells are connected or not, the amplifier has to be put in voltage-clamp mode. Then, a blue laser (473 nm) is used to photostimulate the dLGN axons on the slice, using light pulses of 2 ms and with a power that is not too high so that it elicits unreliable activation of the long range thalamo-cortical axons. In the figure 4.3 D we can see an 8 by 8 grid of stimulating points.

When the light pulse is emitted, multiple axons are stimulated. If among these spiking axons there is at least one that contacts both cells then, across stimulations, the eEPSCs will have a positive correlation, otherwise they

will co-fluctuate independently and their correlation will be close to zero. If we observe that the pair of recorded cells are positively correlated, given a stimulation point, it means that at this location, there is at least one axon coming from the thalamus that contacts both cells.

The fact that the recorded neurons are connected or not doesn't matter for recording eEPSCs in the pairs, because the power of the laser is low enough so that only few dLGN axons spike and the membrane potentials of the recorded pairs will stay always below the spiking threshold, avoiding poly-synaptic responses.



*Figure 4.3: Method to relate local and long range connectivity. A: virus injection in the dLGN. B: slicing of the brain, patching a couple of cells. C: recording the eEPSCs. D: stimulating the projecting axons on a 8 by 8 grid. In each point of the grid, several repetitions are applied. This figure was adapted from Petreanu 2009 [116].*

Before proceeding with the experimental results, we will present the modelling that we made around the main experiment, in which we present how does the membrane potential of the two cells correlate when we vary the power, and also what happens when we stimulate at fixed power. The aim of this modelling is to understand how does the mean correlation depends on the fraction of shared inputs, and how does other biophysical variables impact this mean correlation.

## 4.3 Modelling the covariance of the membrane potential of a couple of patched cells as a function of the proportion of shared inputs

### 4.3.1 Modelling the correlation as a function of the stimulation power

To model the experiment, we will start by simplifying the experimental situation, and we will reintroduce some complexity latter: we will do so by considering certain variables first as constants and then as random variables. Lets consider for a moment that instead of stimulating in 64 different locations (like in figure 4.3 D), we only stimulate in one location, and that in this location there are enough axons from LGN projecting to the pair of recorded cells. Furthermore, lets make the assumption that in the stimulated location, the number of axons that project to only one of the neurons are equal, said otherwise, that the connectivity pattern is symmetric (see figure 4.4). We call  $N$  the total number of LGN axons, and  $r$  the fraction of shared input, i.e., the fraction of axons that contact both patched cells. We define three ensembles of axons ( $E_1, E_2, E_3$ ) that contact respectively only the first cell, both cells, and only the second cell. If in a location there are  $N$  axons, and a proportion  $r$  of shared axons, there are  $N.r$  axons in the central branch  $E_2$  and  $\frac{N.(1-r)}{2}$  in  $E_1$  and  $E_3$ . Implicitly the axons are numbered from 1 to  $N$ , so for example  $i \in E_2$  means  $i \in [1 + \frac{N(1-r)}{2}, \frac{N(1-r)}{2} + N.r]$ .

We define three random variables  $x_1, x_2, x_3$  as the summed number of activated axons ( $x = \sum e_i$ ) in each ensemble of axons. For each axon define a Bernoulli variable  $e_i \in \{0, 1\}$ ,  $i = 1, 2, \dots, N$  that defines a spiking event (spike with probability  $p_i$ ). We model the spiking of an axon as a probabilistic event, because we stimulate in a regime in which failures and successes occur.

Let's call  $\sigma_1, \sigma_2$  the membrane potential in the two cells (expressed in number of excitatory post-synaptic potentials). Because each event  $e_i, i = 1, \dots, N$  happens independently of each other, the variables  $x_1, x_2$  and  $x_3$  are uncorrelated. In addition, there is uncorrelated electrical noise - independent of the laser power - affecting the measure of  $\sigma_1$  and  $\sigma_2$ .

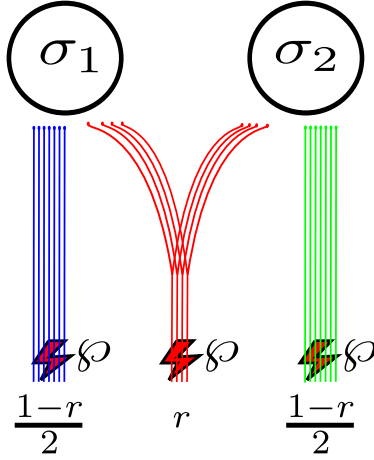


Figure 4.4: Schematic of the model: three ensembles of axons contact symmetrically two recorded cells.  $r$  designates the proportion of shared axons.  $P$  symbolises the power of the laser.

We define this noise as a random variable  $\xi$  of zero mean and variance  $\gamma \ll 1$ .

$$\begin{cases} \sigma_1 = x_1 + x_2 + \xi_1 \\ \sigma_2 = x_2 + x_3 + \xi_2 \end{cases}$$

The events in the central branch contribute to the joint activity of the two neurons (see appendix for a proof) :

$$\text{cov}(\sigma_1, \sigma_2) = \text{var}(x_2)$$

The expected value of each event  $e_i$  is modulated by the power of stimulation  $\wp$  :  $p_i = g(\wp - \theta_i)$ , as observed experimentally [115] (Figure 4.5 A).  $g$  is a sigmoid, a Gaussian cumulative distribution function of zero mean and standard deviation  $\sigma_s$ .  $\theta_i$  is a threshold that is particular to each axon, who obeys a certain unknown distribution  $\text{pdf}_\theta(\wp)$  (figure 4.5 E).

The variance of a spiking event is (see figure 4.5 B).

$$\text{var}(e_i) = p_i(1 - p_i) = g(\wp - \theta_i)(1 - g(\wp - \theta_i))$$

The variance of  $x_2$  can be expressed as :  $\text{var}(x_2) = \sum_{i \in E_2} \text{var}(e_i) = \sum_{i \in E_2} g(\wp - \theta_i)(1 - g(\wp - \theta_i))$  (figure 4.5 D). The variance of the events in

$E_1$  and  $E_3$  doesn't contribute to the covariance of  $\sigma_1$  and  $\sigma_2$  (figure 4.5 D and F).

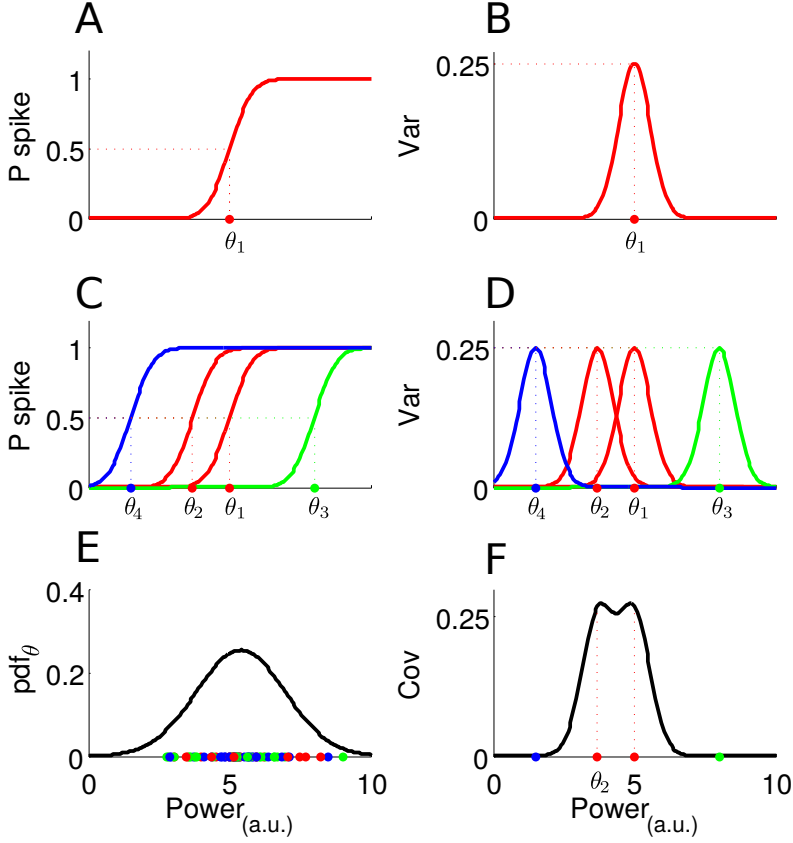


Figure 4.5: A,C: Spike probability of one or many axons. B,D: spiking variability of one or many axons as a function of the laser power. The color code denotes membership to one of the three ensembles  $E_1$  in blue,  $E_2$  in red and  $E_3$  in green. Each dot is one axon threshold. E: probability distribution of the thresholds. F: covariance of  $\sigma_1$  and  $\sigma_2$ .

The correlations can be expressed as:

$$\text{corr}(\sigma_1, \sigma_2) = \frac{\text{var}(x_2)}{\sqrt{\text{var}(x_1) + \text{var}(x_2) + \gamma \sqrt{\text{var}(x_2) + \text{var}(x_3) + \gamma}}}$$

The role of the noise in this model is to have a valid definition of correlation even when the power is zero. Experimentally, this noise is electrical noise; its variance is of the order of one fifth of the variance of the signals. High values of  $var(x_1)$ ,  $var(x_3)$  and  $\gamma$  with respect to  $var(x_2)$  decorrelate the membrane potential of the two neurons.

### 4.3.2 Modelling the mean correlation

When we stimulate at one specific place, there are as much thresholds as there are axons. These thresholds are sampled independently from a continuous distribution. Experimentally, we don't know how this distribution looks like. For illustrative purposes, we chose a Gaussian (4.5 E), but for many of the points that we will make later, the exact shape of this distribution is not so important, it suffices that it is unimodal. In all the following plots we choose two different proportion of shared axons :  $r = 15\%$  (in red) and  $r = 25\%$  (in black).

In figure 4.6 A, we plot the probability distribution function of the thresholds as a function of the laser power. In the abscissa we marked with crosses "+", the thresholds for a particular location site, and we marked with circles the thresholds that were attributed to the axons of the second ensemble of axons  $E_2$ , which projects to both cells. Idem for the black crosses and circles, placed on top of the distribution.

In figure 4.6 B, we plot the correlation as a function of the power, using previously established the formula of the correlation. As we see in the figure 4.5 D and F, only the shared axons (circles) contribute to the correlation or to the covariance.

A very important remark for what will follow in a subsequent paragraph is that at a given power, the correlation can only be big if there are shared axons which have thresholds that are close enough- in a power scale of the order of the length of the variance profile (figure 4.5 D,F), so that they sum up variance : lets recall the formula  $cov(\sigma_1, \sigma_2)_{(\varphi)} = var(x_2)_{(\varphi)} = \sum_{i \in E_2} var(e_i)_{(\varphi)}$ .

In figure 4.6 C, we plotted the expected number of recruited axons  $E_{RA}$ .

$$E_{RA}(\varphi) = E\left(\sum_{i=1}^N e_i(\varphi)\right) = \sum_{i=1}^N p_i(\varphi) = \sum_{i=1}^N g(\theta_i - \varphi)$$

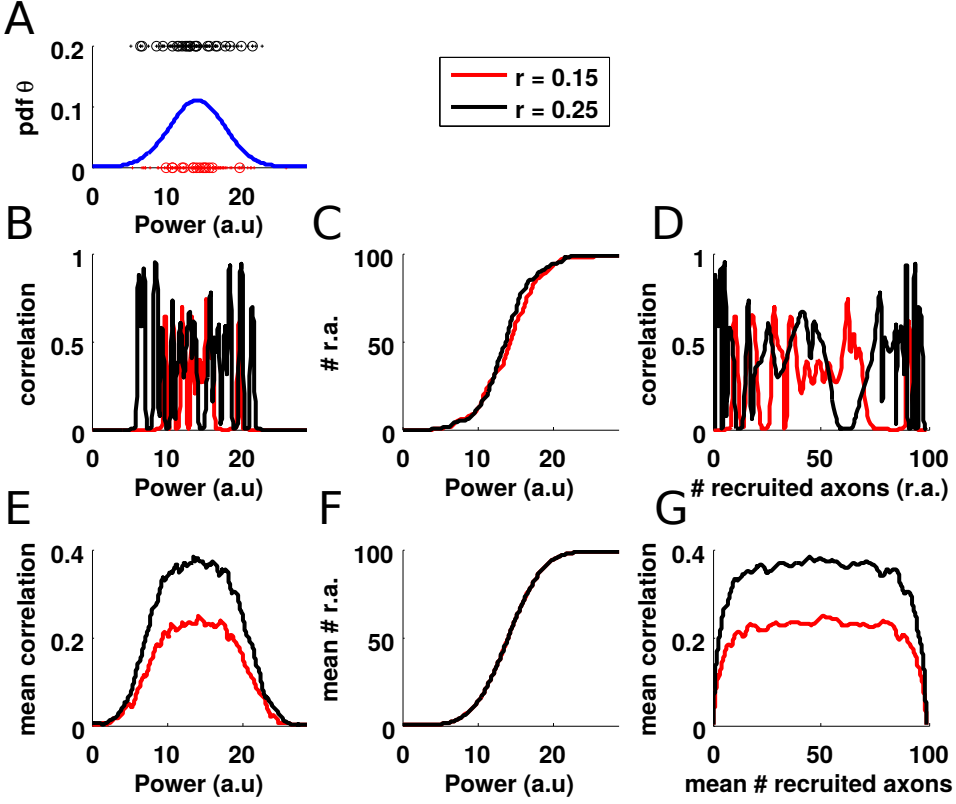


Figure 4.6: Averaging correlations. A: pdf of thresholds, dots: axon sampling for  $r_1 = 0.15$  (red) and  $r_2 = 0.25$  (black). B: Parameters used in this simulation:  $N = 100$ ,  $\text{corr}(\sigma_1, \sigma_2)$  as function of the laser power. Number of recruited axons as a function of the power. C:  $\text{corr}(\sigma_1, \sigma_2)$  as function of number of recruited axons. E: mean correlation as a function of the power. We average many curves like B, each one corresponding to one instantiation of the thresholds. F: idem as in E, but averaging curves like C. G: correlation as a function of the mean number of recruited axons. We obtain this plot, combining the ordinates of E and F.  $\sigma_s = 0, 15$ ,  $\text{pdf}_\theta(\varphi) = N(14, 3.64)$ ,  $\gamma = 0, 0121$ .

Because we sampled the thresholds independently, the red and the black curves are slightly different, but they both converge at high powers to the same quantity:  $N$ .

We could have also plotted the expected number of measured axons,  $E_{MA}$ , which is a quantity that we can measure experimentally.



$$E_{MA}(\varphi) = E\left(\sum_{i \in E_{1,2,3}} e_i(\varphi)\right) = \sum_{i \in E_1} p_i(\varphi) + \sum_{i \in E_2} p_i(\varphi) + \sum_{i \in E_3} p_i(\varphi)$$

In this definition, the shared axons count double, so that at high powers both curves converge to a different quantity that depends precisely on the proportion of shared axons :  $N(1 + r_1)$  and  $N(1 + r_2)$ . We could have used this definition, but on average  $E_{MA} = E_{RA}(1 + r)$ , so they are very similar and the main points we are going to make in this section doesn't depend on which measure we use.

Finally, in 4.6 D, we plotted  $\text{corr}(\sigma_1, \sigma_2) = f(E_{RA})$ , combining the abscissa and the ordinate of B and C.

Now, adding a bit of complexity to the model, lets imagine that we record from different locations, and that in each location, we have a similar connectivity (number of axons, proportion of shared inputs), but with a new sampling in each location from the same threshold distribution. We then average all the obtained correlations and the expected number of recruited axons, at each power (see figure 4.6, E,F,G).

We see that the mean correlation increase with the power and then decreases. This simple fact comes out naturally from the way the variance of Bernoulli variables of probability  $p$  behave as  $p(1 - p)$ : when  $p$  is 0 or 1, the variability is nil, whereas it is maximal at  $p = 1/2$ . This simple fact explains why it is not interesting of doing the experiment at very high or at very low power: because neurons behave deterministically and then they do not covariate.

### 4.3.3 Differences in mean correlation reflect differences in the proportion of shared axons

We are now going to see what can we deduce about the shared connectivity from the correlations. In the previous simulations (figure 4.6 E,G) it was clear that the higher the proportion of shared axons, the bigger the correlations. Indeed, the more axons that belong to  $E_2$  on a single stimulation point can have close thresholds, and then contribute to the correlations: the purpose of this part is to demonstrate this rigorously.

We call kernel  $K$  the bell shaped function that allows to compute the variance of an event, and that we put on top of each particular threshold (see

figure 4.5 B,D). This function takes as parameter  $\sigma_s$ , which determines how steep is the sigmoidal transfer function of stimulation power versus spiking probability, and then how thin is the Kernel (figure 4.5 A,B).

$$K_{\sigma_s}(y) = g(y)(1 - g(y))$$

In the appendix, we show that the variance of  $x_2$  can be written as  $var(x_2)_{(\varphi)} \approx N.r.(K_{\sigma_s} \otimes pdf_{\theta})$ .  $\otimes$  represent a convolution.

The convolution operation between the probability distribution function and the kernel represents well our intuition that, on one side, the recorded neurons covariate at stimulations in which thresholds tend to concentrate because there is more probability mass, and on the other side that the overlap depends on the kernel width.

The variance of  $x_1$  and  $x_3$  can be approximated, using the same argument:  $var(x_1) = var(x_3) \approx N(\frac{1-r}{2})(K_{\sigma_s} \otimes pdf)_{(\varphi)}$  As we have :

$$corr(\sigma_1, \sigma_2) = \frac{var(x_2)}{\sqrt{var(x_1) + var(x_2) + \gamma} \sqrt{var(x_2) + var(x_3) + \gamma}}$$

We will use the notation  $[...]_{\theta}$  to designate averaging the covariance at different stimulation locations. In each location we sample the thresholds from the probability distribution.

Using the previous approximation, we replace it in the expressions of the covariance and obtain:

$$[corr(\sigma_1, \sigma_2)_{(\varphi)}]_{\theta} \approx \frac{N.r.(K_{\sigma_s} \otimes pdf)_{(\varphi)}}{N(r + \frac{1-r}{2})(K_{\sigma_s} \otimes pdf)_{(\varphi)} + \gamma}$$

This has two consequences :

- The first one is that for a given power, if we measure the mean correlation in two populations a and b, that have the same biophysical parameters ( $N$ ,  $pdf_{\theta}$ ,  $\sigma_s$ ) , because the functions  $K_{\sigma_s}(\varphi)$  and  $pdf_{\theta}(\varphi)$  are strictly positive:

$$[corr(\sigma_1, \sigma_2)_{(\varphi)}^a]_{\theta} > [corr(\sigma_1, \sigma_2)_{(\varphi)}^b]_{\theta} \implies r_a > r_b$$

A difference in the mean correlation at a given power reflects a difference in the proportion of shared axons that this neurons receive. This relationship is also valid if instead of power we consider the mean number of recruited

axons  $E_{RA}$  and even the expected number of measured axons  $E_{MA}$ , because of the way the non-linearity expands the abscissa - towards the borders (0 or N)- , when we plot the mean correlations as a function of  $E_{MA}$ , instead of mean correlations as a function of the power.

- The second consequence of the approximation of the mean correlation is that at the maximum of the correlation as a function of the power, where  $var(x_1) + var(x_2) \gg \gamma$  we have:

$$[corr(\sigma_1, \sigma_2)]_\theta \approx \frac{2r}{1+r}$$

The maximal correlation is directly related to the proportion of shared axons. This approximation is accurate for a wide range of parameters and is independent of the biophysical parameters (the threshold probability distribution, the total number of axons and the slope of the sigmoid). In figure 4.7, we see an example of taking the maximal mean correlation and deducing the proportion of shared axons -in dashed lines. In B, we marked with a green dot the true value.

One then can also ask, to which extent is the approximation of the covariance accurate ? Doing simulations we can vary the standard deviation of the distribution  $pdf_\theta$ , and the kernel width  $\sigma_s$  and compute the relative error with respect to the true value  $r$ :  $100 \frac{|\frac{c^*}{2-c^*} - r|}{r}$ . We find (figure 4.7 C) that in a broad range of parameters the approximation is good. The regime of parameters in which the error is maximal is when the standard deviation of the sigmoid is very low and the probability distribution of the thresholds is very broad. This is exactly the case in which, as we work at constant total number of axons, the variance kernels are both too spread apart and too narrow to be added together.

Using this, we could determine the difference in proportion of shared connections for two populations of neurons with similar biophysical characteristics (such as connected neurons and non connected). Let  $c_1^*$  and  $c_2^*$  be the maximal measured correlation of respectively the non connected axons and of the connected axons, inverting the relationship maximal correlation as a function of the proportion of shared inputs we deduce:  $r_1 = \frac{c_1^*}{2-c_1^*}$ ,  $r_2 = \frac{c_2^*}{2-c_2^*}$ , and then  $\Delta r = \frac{c_1^*}{2-c_1^*} - \frac{c_2^*}{2-c_2^*}$ .

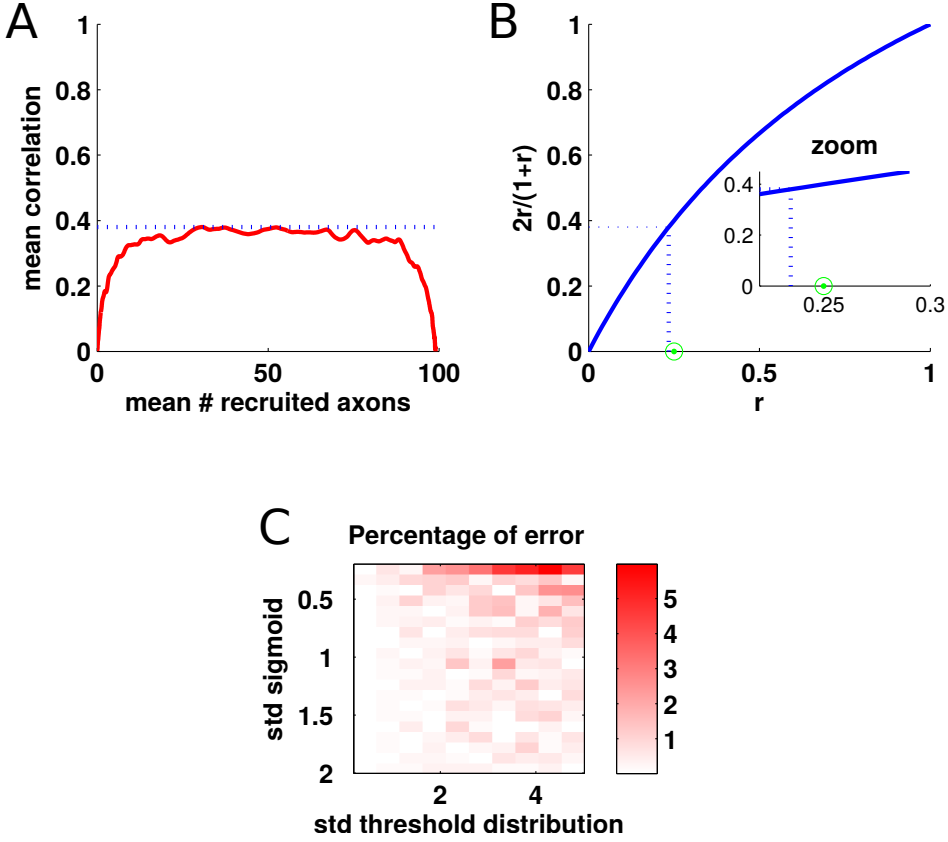


Figure 4.7: Approximation relating the maximal mean correlation with the fraction of shared axons. *A*: mean correlation as a function of the mean number of recruited axons. Dotted line: ordinate of the maximum. *B*: relating the maximum mean correlation with the proportion of shared axons. The true proportion of shared inputs is 25%, which is marked with a green dot in the inset. *C*: computing the error as a function of the standard deviation of the thresholds distribution and of the standard deviation of the sigmoid (transfer function of spiking probability as function of the power). The other parameters used for the simulation, are as in 4.6.

### 4.3.4 Adding variability about the number of recorded sites in one location

In each slice, and for each stimulation site, the number of axons projecting to each neuron  $(n_1, n_2, n_3)$  and the axon thresholds  $\{\theta\}$  are different. Experimentally, applying this same protocol, for given powers one measures the correlations at every stimulation site and then average across sites:  $[[corr(\sigma_1, \sigma_2)_{(\varphi)}]_{\theta}]_{(n_1, n_2, n_3)}$ .

Using numerical simulations, one can show that if we vary randomly the number of axons -and then the proportion of shared axons - in  $E_1, E_2, E_3$ , but keeping the total amount of axons  $N$  in each stimulation site constant, then taking a discrete probability distribution for each axon belonging to each one of the three axon ensembles  $(p_1 = n_1, p_2 = n_2, p_3 = n_3) = (\frac{1-r}{2}, r, \frac{1-r}{2})$ , leads to the equality:

$$[[corr(\sigma_1, \sigma_2)]_{\theta}]_{(n_1, n_2, n_3)} = [corr(\sigma_1, \sigma_2)]_{\theta}^{(n_1, n_2, n_3) = (\frac{1-r}{2}N, rN, \frac{1-r}{2}N)}$$

We then build two models with respectively one or two sources of variability ( $\{\theta\}$  and  $(\{\theta\}, (n_1, n_2, n_3))$ ). We can show that with respect to the quantity that concerns us, the average correlations over these random variables ( $[...]_{\theta}$  and  $[[...]_{\theta}]_{(n_1, n_2, n_3)}$ ), both models are equivalent (see figure 4.8, red and green plots).

For both models, we can also simulate the spiking of all the projecting axons, for every stimulation site, and verify that they correspond well to the quantity predicted by the averaging of the correlation formula see figure 4.8, blue plots).

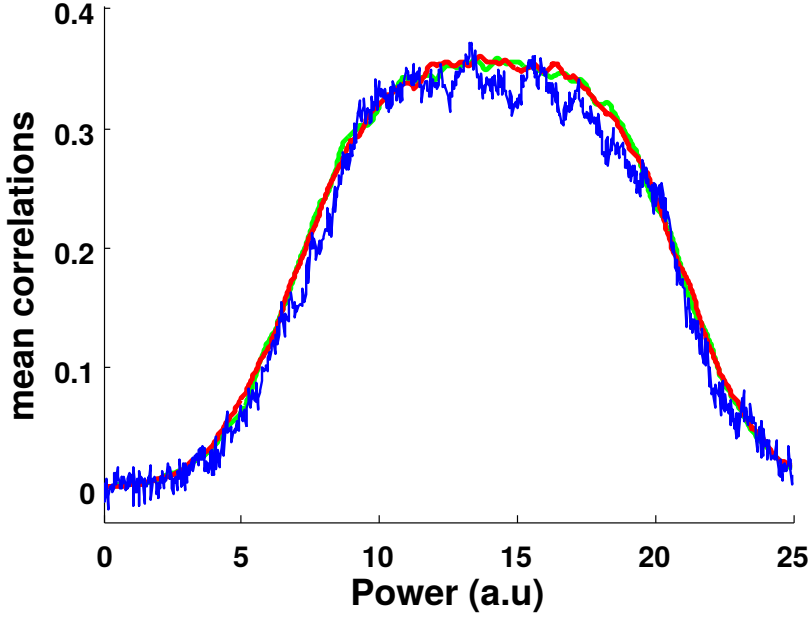


Figure 4.8: Similarity of the correlation model  $[[\dots]_{\theta}]_{(n_1, n_2, n_3)}$  (in red) and  $[\dots]_{\theta}^{(n_1=N \cdot \frac{1-r}{2}, n_2=N \cdot r, n_3=N \cdot \frac{1-r}{2})}$  (in green). The parameters used for the simulation, are as in 4.6. We averaged 5000 different parameter instantiations for both models. In blue we plotted the mean correlation after simulating numerically like if we were stimulating the axons many times (30), in different locations (600). For each location, we sample from the discrete distribution  $(p_1, p_2, p_3) = (\frac{1-r}{2}, r, \frac{1-r}{2})$ , in order to distribute the axons, and the thresholds from the threshold distribution  $\text{pdf}_{\theta}$ .

The following part is a complement to understand what happens qualitatively when we add a layer of variability in the post-synaptic EPSCs.

#### 4.3.5 Examining the consequences of variability in the post synaptic current amplitude

In our model, post synaptic current amplitude is constant, whereas in the data it fluctuates, due to the variation of neurotransmitter quantity in each vesicle. We are going to see that the relevant parameter that intervenes in our model is the variance of the amplitude distribution. This quantity is

easily measurable experimentally, and we are going to see that this variance induces decorrelation of our two cells.

As before, a power elicits or not an event  $e_i$  that traduces the spiking of an axon, but now this event provokes a bigger or smaller amplitude of current in the post synaptic neuron. When there is no event, the amplitude remains at zero, but when there is an event, this amplitude is distributed according to the random variable  $\eta \sim N(1, \omega)$ , which represents the amplitude and which interacts multiplicatively with the events  $e_i$ . The chain of random variables is now:

$$\wp \rightarrow e_i \rightarrow a_i = e_i \cdot \eta_i$$

Let's characterize the random variable  $a$ , thanks to the independence property between  $e$  and  $\eta$ .

$$E(a) = E(e \cdot \eta) = E(e) \cdot E(\eta) = E(e) = 1$$

$$\text{var}(\eta) = \omega^2 = E(\eta^2) - E^2(\eta) \iff E(\eta^2) = \omega^2 + 1$$

$$E(e^2) = p$$

$$E^2(e) = p^2$$

$$\text{var}(a) = \text{var}(e \cdot \eta) = E(e^2)E(\eta^2) - E^2(e)E^2(\eta) = p(\omega^2 + 1) - p^2 = p(1 - p) + p\omega^2$$

As in the previous model :  $x = \sum_i e_i$  is the expected number of recruited axons in a given bundle of axons. Now we are going to define  $A_1 = \sum_{i=1}^r a_i$ , as the summed amplitude of the excitatory post-synaptic currents arriving from the first bundle of axons. In this version of the model we are going to consider a supplementary fact that we didn't considered before, and is that because the central branch of axons splits, there are two axon terminals, so that they are two synaptic vesicles that are released. Before, we considered simply that because the synapses between the LGN and V1 are very reliable [167], every time an axon projecting from the thalamus spiked, an EPSC was evoked, so that when there is a branching in the axonal terminals, there was a constant number of vesicles released at the two sites. In the case in which we consider there is variability in the quantities of neurotransmitter contained in each synaptic vesicle, so that the epscs are more variable, and we consider that in each axon terminal of  $E_2$  ( $A_2^{(1)}$  and  $A_2^{(2)}$ ), a epsc of a different amplitude can happen.

$$\wp \rightarrow e_i \rightarrow a_i^{(1)} = e_i \cdot \eta_i^{(1)}$$

$$\wp \rightarrow e_i \rightarrow a_i^{(2)} = e_i \cdot \eta_i^{(2)}$$

Then the membrane potentials of the two recorded neurons become:

$$\begin{cases} \sigma_1 = A_1 + A_2^{(1)} + \xi_1 \\ \sigma_2 = A_2^{(2)} + A_3 + \xi_2 \end{cases}$$

As in the previous case:

$$\begin{aligned} \text{var}(\sigma_1) &= \text{var}(A_1) + \text{var}(A_2^{(1)}) + \gamma = \sum_{i \in E_{1,2}} \text{var}(a_i) + \gamma \\ &= \sum_{i \in E_{1,2}} p_i(1 - p_i + \omega^2) + \gamma \end{aligned}$$

We have now to recalculate the  $\text{cov}(\sigma_1, \sigma_2)$ , because we have two variables that covary, but that were only one variable in the previous version of the model:  $A_2^{(1)}$  and  $A_2^{(2)}$ . These variables represent the amplitude of summed unitary inputs that come from the second bundle of axons. They both have in common that after a stimulation of the axons, the same axons in the second bundle are on. As this central axon projects to the first and to the second neuron, there are two synaptic releases, that we are going to consider as independent here.

As in the previous model, it is easy to show that :

$$\text{cov}(\sigma_1, \sigma_2) = \text{cov}(A_2^{(1)}, A_2^{(2)}) = \text{var}(x_2)$$

After replacement in the usual formula of the correlation  $\text{corr}(\sigma_1, \sigma_2)$ , we obtain the formula:

$$[\text{corr}(\sigma_1, \sigma_2)]_\theta \approx \frac{N.r.(K_{\sigma_s} \otimes \text{pdf})_{(\emptyset)}}{N(r + \frac{1-r}{2})(K_{\sigma_s} \otimes \text{pdf})_{(\emptyset)} + \omega^2 \sum_{i \in E_{1,2}} [p_i]_\theta + \gamma}$$

We can see that the variability in the EPSCs amplitudes  $\omega^2$  has a decorrelating effect on the mean correlation between the two variables.



## 4.4 Stimulating at only one power and counting the fraction of correlated locations

Experiments are constrained by experimental approaches, and verifying the previous model would take too much time, because we should sample the correlations at many powers, and the cells have a life time duration that only allows for around 30 stimulations.

We then choose to go for a simpler approach experimentally and in the modelling, which is to stimulate at a constant power and observe the fraction of correlated locations as a function of the mean number of recruited axons, and look for differences in the amount of shared input between correlated and non-correlated pairs of cells. The approach consists in comparing across slides of connected and non-connected pairs of neurons, the ratio between the number of stimulated locations eliciting correlated EPSCs over the number of stimulated locations eliciting responses, when we stimulate at constant power. If axons are homogeneously distributed on the slice, this measure is a proxy for the amount of shared axons that pairs of neurons receive.

The strength of a single eEPSCs- that we call the unitary input- is determined by stimulating at low power and observing the minimal amplitude one can obtain across trials. This current trace is integrated in time to obtain a measure of electric charge (in Coulombs C). Also, the mean number of photo-stimulated axons can be approached by taking the mean eEPSC across trials for both neurons, and dividing it by the unitary input. This is just an approximation, because the shared axons count twice, as we discussed in the modelling part when we introduced the quantity  $E_{MA}$ , the expected number of measured axons.

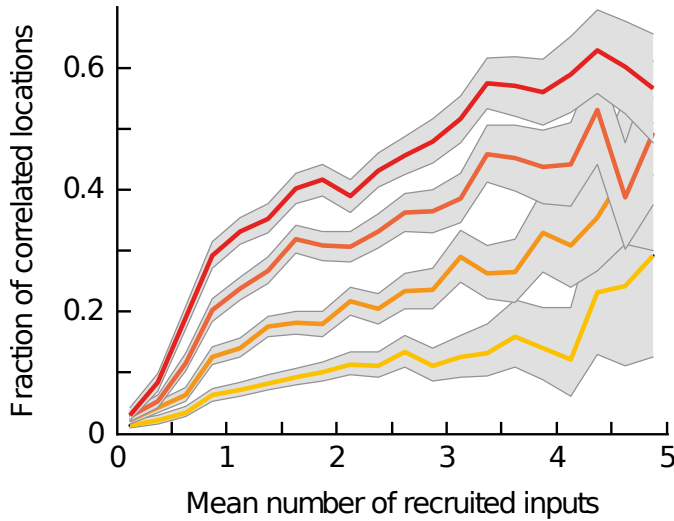
### 4.4.1 Numerical simulation reproducing stimulation at constant power over a grid of points

We performed a simulation in which, at each location, we assigned new thresholds at a variable number of axons, and we stimulated many times at constant power, to obtain correlation in each point, similarly to what was done in the figure 4.8. However, instead of computing the mean number of recruited axons at a given power across locations, we regrouped the mean number of recruited axons by their number, and averaged their correspond-

ing correlations. The simulation showed that the mean correlation increased with the mean number of recruited axons. The bigger the fraction of shared inputs, the bigger the plot of mean correlations as a function of the mean number of recruited axons.

Similarly, we can plot the fraction of correlated locations as a function of the mean number of recruited axons, which is the way the experimental data is plotted (see figure 4.9 and figure 4.10).

To determine whether the correlations are statistically significant at one location, we have to compare the correlation obtained with the distribution of correlations that we obtain when we permute the temporal order of the events for each cell independently, destroying the covariance. If the data is in the top 5% of the correlation distribution, we consider that the observed correlation has few chances (5%) of having being generated by chance.



*Figure 4.9: Simulation of the fraction of correlated locations as a function of the mean number of recruited axons, for a shared fraction of inputs of 10% (yellow), 20%, 30% and 40% (red). Figure adapted from [104].*

The modelling we did in the beginning of this chapter, shows analytically that the recipe to compare at a single stimulation power is to average all the correlations across locations for connected and for non-connected pairs of cells. In this simulation, we do not do that: the axons that spike are the

ones whose threshold is smaller or equal to the laser power. Across locations, the majority of locations have a certain mean number of recruited axons (like 0 or 1), and from that on, the mean number of recruited axons for a given stimulation decreases smoothly, as we see that it also happens in the data (figure 4.10). However, because all axons sample the thresholds from the same distribution, the higher the number of recruited axons, the higher the number of axons belong to  $E_2$  and then the higher the mean correlations.

### 4.4.2 Experimental results

Figure 4.10 shows the fraction of correlated locations pulling together the slides in which cells were connected and the slides in which cells were not connected. The three plots show whether the pairs are in L4 or in layer 2/3 or if alternatively one of the patched cells is in L2/3 and the other in layer 4. As we see in the upper inset, for a low mean number of inputs, we have more data points to establish a difference between the connected and the unconnected pairs.

In figure 4.11 A,B,C, we proceed to sort the data according to whether the fraction of correlated pairs were connected or not. As in figure 4.10, we plotted the cells as a function of the mean number of inputs.

In figure 4.10, we can see, as supported by the simulation in figure 4.9, that a difference in the fraction of shared inputs elicits a quicker growth in the number of correlated locations as a function of the mean number of recruited inputs. We then see how for pairs of cells in layer four (figure 4.10 B), and in simultaneously recorded from L4 and L2/3 (figure 4.10 C), there is a difference in the growth speed. We compute the statistical significance, pulling together the data of one and two mean number of inputs, for which there is more data (in grey). We show in the bottom row, that this difference is statistically significant for L4-L4 pairs and L4-L2/3 pairs (E,F). In the case, of L2/3- L2/3 pairs the difference in growth is not statistically significant (A,D).

Cossell and colleagues [29] showed recently that although cells in L2/3 received many inputs from other cells in L2/3, the ones that were important for predicting the receptive field of the cells were the very strong ones, which were very few. In this study, only weakly connected pairs in L2/3 were measured, (see supplementary material [104]). If the thalamic projections onto L2/3 follow the same logic as the thalamic projections onto L4, and also

of the L4 to L2/3 projections, then the prediction for the thalamocortical projection onto L2/3 is that only strongly connected pairs in L2/3 receive more shared input from the thalamus.

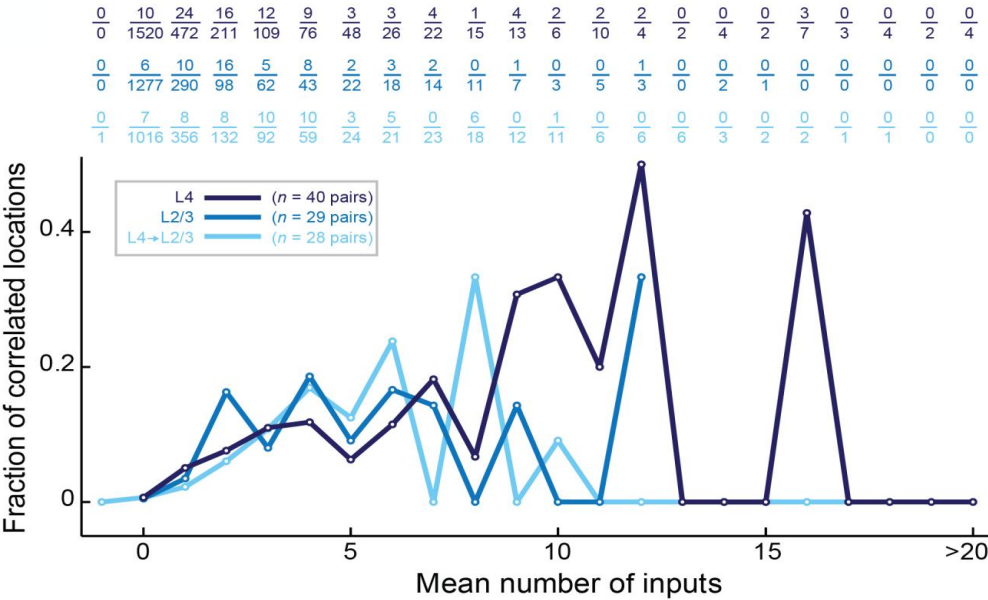


Figure 4.10: Fraction of correlated locations as a function of the mean number of inputs. In the top panel, we can see the measured number of correlated locations over the total number of photo-stimulated locations for L4, L2/3 and L4-L2/3. The pairs of cells that are connected and non connected are merged. Figure adapted from [104].

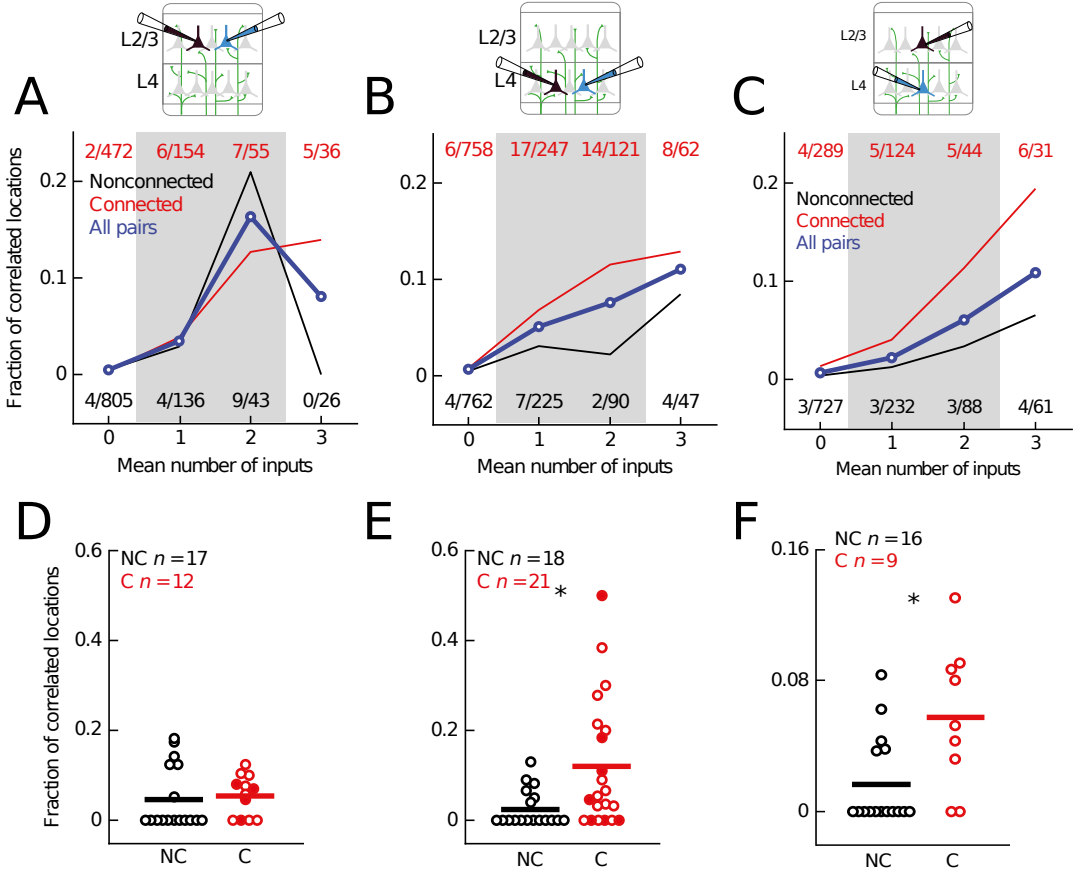


Figure 4.11: Measured number of correlated locations as function of the mean number of inputs, for pairs in L2/3 (A), L4 (B), and L4-L2/3 (C). We concentrate on the first mean number of inputs (0, 1, 2, 3), for which there are more recruited cells, and separate the fraction of recruited axons depending on whether they are connected (in red) or not (in black). In the bottom row (D,E,F), we consider only the fraction of correlated locations corresponding to one or two mean inputs. E and F are statistically significant, while D is not. Figure adapted from [104].

## 4.5 Discussion

### 4.5.1 Summary

The main findings of the study can then be summarised as following:

- dLGN projections innervate strongly -and preferentially- layer 4 and deep L2/3.
- Pairs of connected cells tend to have more incoming shared projecting axons from the thalamus, than cells that are not connected in L4. Idem when we consider pairs of cells such that one of the two is in layer 4, and the other one is in L2/3.
- In pairs of cells recorded in L2/3, there is not such a tendency. Pairs of cells that are connected in layer 2/3 doesn't seem to receive more shared input from the thalamus than pairs of non connected cells.
- This study confirmed that in visual cortex, like it has been shown in other sensory cortices, the thalamo-cortical projection targets also L2/3 in addition of L4. Based on the current findings, it appears also that the thalamocortical axons excite simultaneously pairs of neurons on L2/3 and L4 which themselves are connected.

### 4.5.2 Modelling correlations

From the first part of the modelling study, we obtained a good understanding of the way that biophysical details impacted on the mean correlation across pairs, trials and locations. With this model, one can do three things:

- 1- Sample the axons at different powers and for each power, determine the mean correlation, and plot mean correlation versus power, or alternatively mean correlation versus expected number of measured axons as in [4.6](#).
- 2- At the maximum of the previous plot measure the mean correlation  $c^*$  and deduce the fraction of shared input  $r = \frac{c^*}{2-c^*}$ .
- 3- Repeat the last two steps for another population with similar biophysical characteristics, (if the previous measured population was connected cells, do it for example with non-connected cells), and compare the plots to see if the fraction of shared inputs are different.

Regarding the more experiment related simulations, we considered grids in which in each stimulation site, we affected a different number of axons and different thresholds. We simulated trials with many stimulations and

computed the correlations for each recording site and determined whether this correlations were statistically significant. As in the data, the model shows that the fraction of correlated inputs increases as a function of the number of inputs recruited at each location.

### 4.5.3 Connectivity and function

Since the pioneering works of Santiago Ramón y Cajal, scientists like David Marr have tried to use the anatomy to constrain the range of possible computations that a neural circuit might be performing.

Recently there has been a burst of enthusiasm for what is called the connectome, which is the mapping of all the possible connections in the nervous system of an animal. There are many techniques that could be used to perform this connectome. Each of them has a different resolution in the level of biophysical detail it can grasp.

Many of the must cutting-edge tools are available to trace the connections: genetically modified viruses, opto-genetics, calcium indicators, optics, patch-clamp, dense electron microscopy reconstruction, and even sparse coding approaches to reconstruct connection strengths [71].

The techniques and the word "connectome" are new, but the idea of tracing the connections is old. The cortex has concentrated the joint efforts of electro-physiologists for decades to try to understand the rules that govern the connectivity between layers and between cell types. The hope is that, despite is enormous complexity, the stereotypy of the canonical microcircuit (figure 1.1), will reveal fundamental insights about the computations it might be performing.

However, deducing the dynamics from the connectivity is a very hard problem. Eve Marder, with its forty years experience of studying central pattern generators in the somato-gastric system states: "the connectivity diagram, or connectome, is absolutely necessary and completely insufficient to explain circuit dynamics". After probing the existence or the absence of all the connections in the somato-gastric nervous system -of a few hundred cells-, in order to be able to recover using simulations the dynamics of the central pattern generators, Marder and collaborators had to add many more biophysical details like the connectivity strengths and the ion channels dynamics of different cell types.

In the modelling chapter, we tried to make the link between the connec-

tivity and the dynamics of a neural circuit, and we could understand from a modelling point of view, how quantitative changes in the strength of the connectivity might lead to changes in the dynamics. We can also think, given the modelling we have done, that the specificity of the connections of certain group of cells with another group of cells is crucial to understand how one circuit constrains its computations. Sampling the connections might not be sufficient to grasp the circuit specificities. However, from an experimental point of view, already mapping all the connections -their existence, not even the strengths- of a portion of tissue might be already a huge technical challenge.

Some very promising techniques, like electron-microscopy coupled with machine learning algorithms, are being extensively used in order to reconstruct all the connections from all the neurons in a piece of tissue. For example, the initiative called "EyeWire" aims at reconstruct the whole bodies of retinal neurons and their connections, and is played online by thousands of people. The advantage of this approach, with respect with patch-clamp, is that it might lead to identify all the connections in a tissue, so that the numbers will be more robust, because all the synapses will be counted in a single stack of tissue rather than being sampled from different animals, as it is done in patch-clamping.

However, the major drawback of the electron microscopy approach, is that it doesn't give an estimate of the connectivity strength of synapses. As we just saw, having only the information about whether one neuron is connected with another might not be sufficient at all to say something about the computation that a recurrent circuit might be performing. To nuance this claim, let's recall that Douglas and colleagues [38] had a deep insight about the cortical amplification from signals coming from the thalamus using only data of 3D reconstructed cell bodies and axons using electron microscopy.

Another minor drawback of whole cell body reconstruction using electron microscopy is that for now, in spite of the gigantic efforts and the industrial scale at which the "EyeWire" project is being done, the pieces of tissues that are being studied are small, which precisely might prevent from identifying the long range axons, which are the ones we are interested in.



#### 4.5.4 Importance of long range projections to understand cortical functions

Primary cortical areas receive all kinds of long range projections -which by definition are not primary sensory inputs. Those long range projections may be: motor related inputs, feedback from higher cortical areas or projecting axons from other sensory modalities. Also, these long range projections may be neuro-modulatory and the signals conveyed, be related to attention, expectation or reward processes.

The existing techniques to map long range connections are genetic tools like rabies virus, which jump to the pre-synaptic cells of the infected cells and express channel-rhodopsin in them.

However, patch clamping is the only technique that allows to measure at the same time the existence of connections between two neurons, the synaptic sub-threshold currents that cross their membrane and the connectivity strength between two cells. It has led to numerous important insights about the inputs a cell receives, as we saw in the case of Lien and Scanziani [87]. The presented technique deals with the interaction of local and long range connectivity, and could be applied to any other long range projection.

## 4.6 Appendix

### 4.6.1 Mathematical proofs

#### 4.6.1.1 Covariance of the two neurons membrane potential as a function of the variance of the axons total activity

We will prove the following equality:  $cov(\sigma_1, \sigma_2) = var(x_2)$ .

$$\begin{aligned} E(\sigma_1 \sigma_2) &= E((x_1 + x_2 + \xi_1)(x_2 + x_3 + \xi_2)) \\ &= E(x_1 x_2 + x_1 x_3 + x_2^2 + x_2 x_3 + \xi_1(x_2 + x_3) + \xi_2(x_1 + x_2) + \xi_1 \xi_2) \\ &= E(x_1)E(x_2) + E(x_1)E(x_3) + E(x_3)E(x_2) + E(x_2^2) + 0 \end{aligned}$$

$$E(\sigma_1)E(\sigma_2) = E(x_1)E(x_2) + E(x_1)E(x_3) + E(x_3)E(x_2) + E^2(x_2)$$

$$cov(\sigma_1, \sigma_2) = E(\sigma_1\sigma_2) - E(\sigma_1)E(\sigma_2) = E(x_2^2) - E^2(x_2) = var(x_2)$$

#### 4.6.1.2 Approximation of the variance

We will prove the following approximation:  $var(x_2)_\wp \approx N.r.(K_{\sigma_s} \otimes pdf_\theta)$ .

$$\begin{aligned} var(x_2)_\wp &= \sum_{i \in E_2} K_{\sigma_s}(\wp - \theta_i) \\ &= \sum_{i \in E_2} \int_{-\infty}^{+\infty} K_{\sigma_s}(\wp - \theta) \delta(\theta - \theta_i). d\theta \\ &= \int_{-\infty}^{+\infty} K_{\sigma_s}(\wp - \theta). \left( \sum_{i \in E_2} \delta(\theta - \theta_i) \right). d\theta \\ &\approx \int_{-\infty}^{+\infty} K_{\sigma_s}(\wp - \theta). (N.r.pdf(\theta)). d\theta = N.r.(K_{\sigma_s} \otimes pdf)_{(\wp)} \end{aligned}$$

#### 4.6.1.3 Covariance of the two neurons membrane potential when there is variability in the post synaptic current amplitude

We will prove the following equality:  $cov(\sigma_1, \sigma_2) = cov(A_2^{(1)}, A_2^{(2)}) = var(x_2)$

$$\begin{aligned}
 cov(A_2^{(1)}, A_2^{(2)}) &= E(A_2^{(1)} A_2^{(2)}) - E(A_2^{(1)}) E(A_2^{(2)}) \\
 &= E\left(\left(\sum_{i=1}^r e_i \cdot \eta_i^{(1)}\right) \left(\sum_{i=1}^r e_i \cdot \eta_i^{(2)}\right)\right) - E\left(\sum_{i=1}^r e_i \cdot \eta_i^{(1)}\right) E\left(\sum_{i=1}^r e_i \cdot \eta_i^{(2)}\right) \\
 &= E\left(\sum_{i=1}^r e_i^2 \eta_i^{(1)} \eta_i^{(2)}\right) - E^2\left(\sum_{i=1}^r e_i\right) \\
 &= \sum_{i=1}^r E(e_i^2) E(\eta_i^{(1)} \eta_i^{(2)}) - E^2\left(\sum_{i=1}^r e_i\right) \\
 &= \sum_{i=1}^r E(e_i^2) E(\eta_i^{(1)}) E(\eta_i^{(2)}) - E^2\left(\sum_{i=1}^r e_i\right) \\
 &= \sum_{i=1}^r E(e_i^2) - E^2\left(\sum_{i=1}^r e_i\right) = var(x_2)
 \end{aligned}$$

### 4.6.2 Numerical simulation

Experimentally, at any given photostimulation location, the number of axons on each ensemble ( $n_1, n_2, n_3$ ) and the axonal thresholds  $\{\theta\}$  are different. For the plot in figure 4.9, we simulated sampling axons from different locations with fixed laser power. Axons were randomly assigned to the ensembles  $E_1$ ,  $E_2$  or  $E_3$ , with probabilities  $(\frac{1-r}{2}, r, \frac{1-r}{2})$  respectively. The number of axons was fixed at 40. Each axon had a threshold  $\theta$  sampled from a Gaussian probability distribution with mean 20 (a.u. of power) and standard deviation 5. Laser power was fixed at 10 a.u. The variance of the noise,  $\gamma$ , was 0.2. The sigmoid  $g$  was a cumulative Gaussian distribution centered at zero and with a standard deviation of 2. The tested fractions of shared axons  $r$  were 0.1, 0.2, 0.3 and 0.4. We obtained values of the correlation coefficient  $corr(\sigma_1, \sigma_2)$  after 20 trials. By bootstrapping 10,000 times we estimated 99.5 % confidence intervals around the value of  $corr(\sigma_1, \sigma_2)$ . We considered the simulated locations correlated when the lower boundaries of the confidence intervals were greater than zero. The number of evoked inputs per trial  $I$  corresponds to the number of spiking axons, where shared axons count double:  $I = \sum_{i \in E_1} e_i + \sum_{i \in E_3} e_i + 2 \sum_{i \in E_2} e_i$ .



# Chapter 5

## Determination of the number of statistically significant principal components

### HIGHLIGHTS

- We start this chapter by presenting a brief overview of some of the existing methods to determine the number of statistically significant principal components.
- We then introduces one new method using non-parametric statistics.
- We compare how well does our method performs with respect to the other ones.

### Contents

<b>5.1</b>	<b>Introduction</b>	<b>214</b>
<b>5.2</b>	<b>State of the art</b>	<b>217</b>
5.2.1	Parallel analysis	218
5.2.2	MAP test	221
5.2.3	Cross-validation methods	222

- 5.3 Proposed method . . . . . 225**
  - 5.3.1 Motivations . . . . . 225
  - 5.3.2 Preprocessing . . . . . 225
  - 5.3.3 Shuffling the data: preserve the variance, destroy the covariance . . . . . 226
  - 5.3.4 Use of this shuffle to determine the statistical significance of the principal components of the data 227
  - 5.3.5 Signal and noise . . . . . 229
  - 5.3.6 Detailed procedure . . . . . 231
  - 5.3.7 Explanation . . . . . 232
- 5.4 Comparing different methods . . . . . 236**
  - 5.4.1 Comparing methods with different data-bases . 236
  - 5.4.2 Comparing methods with different artificially generated data sets . . . . . 238
- 5.5 Determining the different number of statistically significant components in different experiments during desynchronised state . . . . . 241**
- 5.6 Discussion and conclusions . . . . . 241**
  - 5.6.1 Results from the comparison with simulated data 241
  - 5.6.2 Results from the analysis of spiking data from desynchronised state . . . . . 242
  - 5.6.3 Advantage of the preprocessing to evaluate non-stationarity of the data . . . . . 242

---

5.1 Introduction

Principal component analysis (PCA) is the oldest and the most popular method of multivariate analysis. This technique was discovered by Pearson in 1901 [112], and rediscovered though a different method by Hotelling in 1933 [70]. The core idea behind PCA is to observe the data from another perspective, trying to identify common patterns of variation in the observed variables. More precisely, if the jointly observed variables are generated by

a multivariate stochastic process, PCA is going to try decorrelate this processes though a rigid rotation of the canonical basis into another orthonormal basis (figure 5.1, A and B).

The dataset of jointly observed  $N$  variables is composed of an observation matrix  $X$  of dimensions  $T$  times  $N$ ,  $T$  is the number of observations. We center the observations with respect with their mean, in order to see common variations around the mean, and we usually normalize the variables by their standard deviation. If we call  $C$  the covariance or the correlation matrix,  $C = \frac{1}{T}X'X$ . PCA consists simply in diagonalizing this matrix:

$$C = P.D.P'$$

$D$  is a matrix whose diagonal elements  $\lambda_i$  are the projected variances of the multivariate data  $\lambda_i$  onto the column unitary vectors  $PC_i$  of the rotation matrix  $P$ : these vectors  $PC_i$  are called the principal components. We usually sort the diagonal matrix from the highest to the lowest variance, so that when we refer to the first principal component, we always refer to the direction that explains more common variance in the data.

Said otherwise, PCA performs decorrelation through orthogonalization. This decorrelation works when the data we are dealing with has an ellipsoidal shape in high dimensions (ex. figure 5.1 A). We can see in figure 5.1 B, that this method doesn't always work: even if the first PC captures the direction with the highest projected variance, and the second PC does it with the constraint of being orthogonal to the first  $PC$ , we can see that the data in those axis is not independent, but rather linked by a non-linear relationship.

If there is a gap in the amount of variance explained by a first subset of principal components with respect to the rest of the components, i.e. if these subset of variables seem to explain much more variance, as data miners we can postulate that there are many independent sources of variability in the data, and even call **signal** the processes that happen on those PCs (see red in figure 5.1 C) and **noise** the processes that happen on the axis that have less variance (in blue in the bottom figure).

This study is concerned precisely with how to determine in a principled way the number of principal components which are statistically significant, and then call them signal. As figure 5.1 C shows, PCA is a bijective linear transformation from  $R^N$  to  $R^N$ , it is just a basis rotation. Per se, it doesn't tells us what is signal and what is noise.

However, the common usage of PCA -which is not the result of PCA- is

to reduce the dimensionality of the data set, retaining only the first PC's, to simplify the interpretation, while retaining as much as possible of the explained variance. For example if the data is unidimensional, we will find that a linear combination of the variables, explains a good part of the variance. The experimenter might then look for interesting interpretations of the covariation of the variables, postulating the existence of a hidden underlying variable that correlates the observed variables: if for example the only measured variables are weight and size, one hidden variable that might explain much of the common variance is the age.

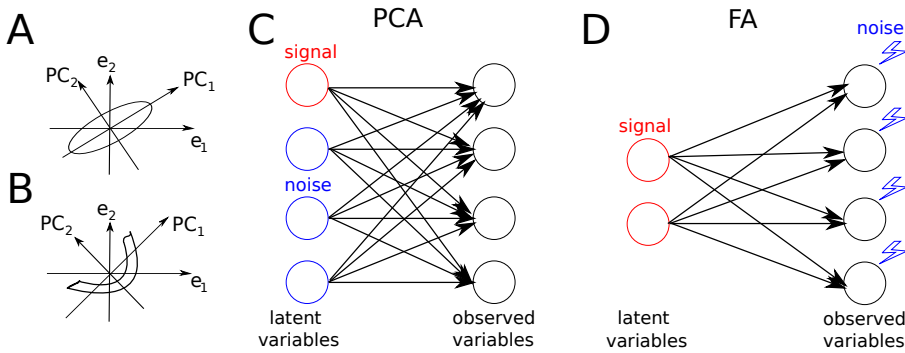


Figure 5.1: Description of PCA and comparison with FA, as they are similar tools for exploratory data analysis. A: Example of data for which PCA performs well. B: example of data in which PCA fails to decorrelate the data in the new axis. C: PCA seen as a linear combination of latent variables, some of which will be called signal, and some of which will be called noise. D: FA seen as a linear combination of latent variables. Each observed variable has independent noise.

What we call noise is then considered to be independent variance attributable to individual variables or to subsets of elements. However, whereas the definition of what is signal and how to observe it is more or less clear in PCA, the interpretability of the noise is less clear, because the orthogonal complement of a given "signal" subspace doesn't points necessarily in the direction of individual neurons, so the noise is still usually a linear combination of the latent variables.

In **factor analysis** (FA) we have -as in PCA- a certain number of latent variables that combine linearly to produce observed variables, but the difference is that in FA these observed variables can have independent sources



of noise (figure 5.1 D). The advantage of PCA with respect to FA is that is simpler to use and to interpret. For example, in PCA we can have access to the scores. In FA, as opposed to PCA, the factors (equivalent of the PC's) are not unique, and this severely constraints the interpretation one can draw from these factors [15]. Overall, these two methods are very similar and complementary, and in both we have to make the choice of how many latent variables we are going to retain.

Even if the problem that we will be dealing with in this study is mostly technical, its repercussions are important due to the widespread of the use of PCA all around quantitative disciplines. In neuroscience, with the introduction of parallel recording techniques (like electrodes, imaging, eeg,...) as well as the acceleration of computing power, we can start asking questions about how do neurons perform computations at the circuit level, or in simultaneous in different brain areas. For example Shenoy, Sahani, and colleagues have been looking at electro-physiological data with the idea of finding low dimensional trajectories that can simplify the highly complex picture of hundred of cells recorded simultaneously and correlating these trajectories with behaviour. They have also developed more advanced methods to extract latent variables that are more adapted for spiking data like GPFA [165] and GLDS [93].

## 5.2 State of the art

There are dozens of methods that have been proposed to estimate the number of factors both for PCA and for factor analysis. In factor analysis the problem is even more critical than in PCA, because it leads to a distorted view of the data through the fuse or splitting of factors and to excessive loadings. In PCA these considerations doesn't apply, because PCA is unique. Once the number of factors is determined, we are faced in both methods with the problem of interpreting those factors.

Many methods have been proposed to solve the problem of the number of factors. Some where purely empirical rules like the scree test (Cattel, 1966) [24] Some others like Bartlett's rule (1950) [9] are statistically based methods, but which assume normality. A couple of these early statistical methods had a more lasting impact and are still very used in statistical software packages like SPSS -mostly Paralell analysis from Horn (1965) [69],

and Minimum Average Partial from Velicer (1976) [152].

More recent methods for determining the number of principal components, based on cross-validation, [19], [36] have made a significant impact on the statistical field, even if they still haven't reached a broader audience of practitioners, who mostly uses software packages or simple recipes like Kaiser's rule [75] or the scree test [24].

We will present the most relevant methods found in the literature. We will start with the oldest methods of Horn and Velicer called respectively Parallel analysis and Minimum Average Partial test, and we will pursue with the more recent state of the art methods based on cross-validation.

### 5.2.1 Parallel analysis

Horn [69] argues in that if the estimated correlation C matrix is determined on N random variables, it is expected to have N non zero eigenvalues all nearly equal to 1. However, because of the sample finite size, the eigenspectrum is not a flat line with all the values equal to 1, but a curve "with slope indicating the extent to which sampling error and least-squares bias have combined to increase the value of the correlations". In figure 5.2 we simulated sampling from a multivariate uncorrelated Gaussian distribution at two different sizes.

When we average the eigenspectrum over many realisations of the covariance matrix, at the point  $N/2$ , for symmetry reasons, the curve is expected to cross the abscissa 1. Horn cites Guttman (1954) [56], which established an inequality to estimate the maximum rank  $r$  of a correlation matrix:  $r \leq \text{cardinal}(\{\lambda | \lambda \geq 1\})$ . This upper bound -also called Kaiser's rule- is simply the number of eigenvalues of the correlation matrix that are larger than one. Horn decides to take the Guttman upper bound on the number of factors into account, but after discounting for the size effect on the shape of the eigenspectrum, because it was not taken into account in Guttman's demonstration: "This proof is based on an assumption that there is no error due to sampling in the population of subjects, that sampling takes place only in the universe of measures (tests)" [56].

The accurate way to compare the significance of the data is, according to Horn, to compute the eigenspectrum over many shuffles and then to average it, eigenvalue by eigenvalue. The PC for which the eigenspectrum of the data is above the average shuffle curve, is significant (see figure 5.3).

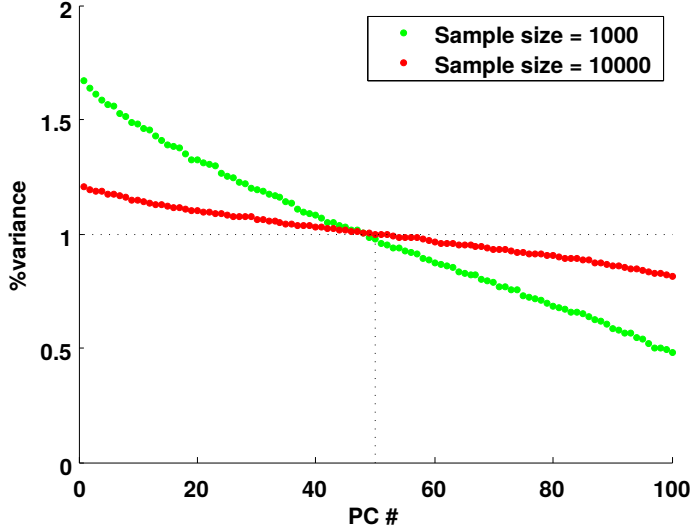
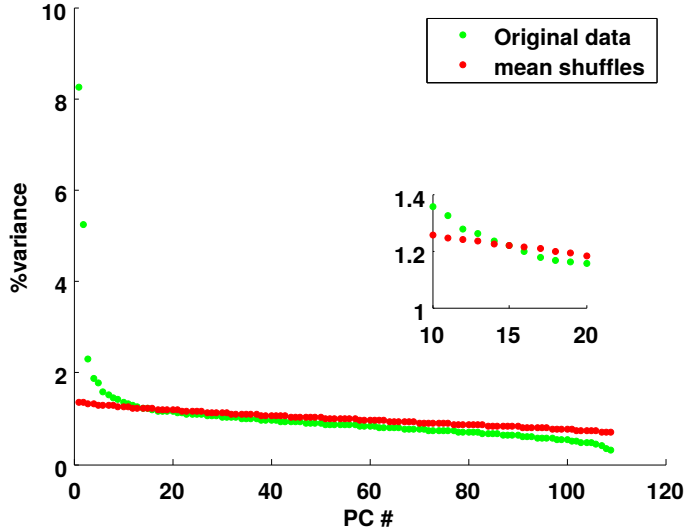


Figure 5.2: Effect of the sample size on the distribution of the eigenvalues. Both sets of samples were generated from a multivariate Gaussian distribution of uncorrelated variables, of zero mean and of unitary variance.

Parallel analysis is an adaptation of the Kaiser's rule that takes into account the finite size effect of the sampling. It introduces a very interesting feature which is to create shuffles and to compare the eigenvalue of the data with the distribution of shuffled data. However, independently of the quality of Parallel Analysis to estimate accurately the number of significant components, we must recall that this method is not doing hypothesis testing, in the sense that the surrogate data is not being used to generate the null hypothesis, which is the following: the data is statistically undistinguishable from the shuffled data. In the shuffled data, the correlations between variables are destroyed by random permutations of the temporal occurrences, but remain finite due to finite size effects.

If we considered precisely using shuffles as in Parallel Analysis into an hypothesis testing framework, we would like to see for example if when we shuffle the data, and create a distribution that is spherical on average, the first eigenvalue of the data lies into the distribution of projected variances of shuffled data onto the  $PC_1$  of the original data.

Some more recent versions of parallel analysis (see [63] for a review),



*Figure 5.3: Example of parallel analysis applied to real data. We performed 1000 shuffles, and computed every time the eigenspectrum of the covariance matrix obtained, and then averaged those eigenspectrums. Inset: zoom of the eigenspectra between the components 10 and 20. According to parallel analysis, this data set has around 14 components that are statistically significant.*

consider rather to perform  $Q$  shuffles on the data, and compute the intervals encompassing [5%, 95%] of the distribution of each eigenvalue, and then consider as significant the components that are above the upper significance boundary.

This method goes in the right direction, but still runs into some problems. For instance, the shuffling preserves the marginal variances, but the marginal variances are a function of both the signal and the noise. Thus, variance from the signal is spread into all directions, and the dimensions with lower variance appear as to have systematically less variance than the eigenspectrum of the shuffled version.

We will show that this approach goes in the right direction, but that it still runs into some problems, and we will propose a method that overcomes them. Before doing so, we will first some other important methods to asses the number of principal components. We will start by a method called Velicer's Minimum Average Partial test or simply MAP test.

### 5.2.2 MAP test

In 1976, Velicer came out with a very clear and elegant way of determining the number of components using the partial correlations [152].

We know that, given that a correlation matrix is symmetric and real, its eigenvectors and eigenvalues are real, and moreover, its eigenvectors are orthogonal. Diagonalizing the correlation matrix leads to the following expression (see appendix 3.10.4 for a demonstration):

$$C = P.D.P^{-1} = P.D.P^t$$

By developing the last expression, we obtain the spectral decomposition of  $C$ :

$$C = \sum_{i=1}^N \lambda_i P C_i . P C_i'$$

Now lets image that we have only two observed variables X and Y, which are a linear combination of two latent variables Z and W. The partial correlation quantifies how much does one latent variable Z contributes to the total correlation between the two observed variables by removing the influence it has on X and Y. The partial correlation between X and Y, controlling for Z is :

$$\rho_{XY.Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{1 - \rho_{XZ}^2}\sqrt{1 - \rho_{YZ}^2}}$$

Velicer reminds us that we can compute easily the matrix of partial correlations after having removed k latent variables by renormalizing the matrix so that it is a correlation matrix.

$$C_k = C - \sum_{i=1}^k \lambda_i P C_i . P C_i'$$

$$C_k^* = G . C_k . G \quad G_{ij} = \frac{\delta_{ij}}{\sqrt{(C_k)_{ij}}} \quad \delta_{ii} = 1, \delta_{i,j \neq i} = 0$$

Looking when does the mean square partial correlations attains a minimum tells us which are the ideal number  $k^*$  of components.

$$f_m(k) = \frac{1}{N(N-1)} \sum_i \sum_{j \neq i} (C_{ij}^*)^2$$

In his revised version of this test in 2000 [153], Velicer proposes a very similar statistic, but elevating the partial correlations to the fourth power:

$$f_m(k) = \frac{1}{N(N-1)} \sum_i \sum_{j \neq i} (C_{ij}^*)^4$$

The explanation given by Velicer of such increase and decrease of such features is the following: the formula for the partial correlation  $\rho_{XY.Z}$  of two variables X and Y, controlling for a third variable Z,  $\rho_{XY.Z}$  decreases if the numerator decreases faster than the denominator, but increases in the opposite case. For example, if  $\rho_{XZ}$  is big and  $\rho_{YZ}$  is small, then it means that the latent variable Z is only driving one variable, and not driving common variance.

### 5.2.3 Cross-validation methods

Cross-validation is a very useful statistical tool to perform **model selection**. The main idea is to cut the data in two non-overlapping parts, one of them called the **training set** and the other the **test set**. In the training set do a series of model fits of increasing complexity that decrease the error between the data and the model **prediction error** on the training set. However, in order to see to which extent the statistical regularities captured by the model on the training set are also present in the test set (how well these models **generalize**), we compute the prediction error of each of these models in the test set, and we keep the model that has the lower error.

A very simple case is when we have a cloud of points representing an independent variable X and a dependent variable Y. Omitting a point at a time, to have a training and a test set, we can fit a polynomial of various orders, and the higher the order of the polynomial, the lower the prediction error on the training set.

In this way, as the complexity of the polynomial increases we will see the mean square error decrease towards a minimum and then re-increase. This minimum gives the good trade-off between bias and variance that captures as much variance as possible, but that doesn't over-fit the data, so that it generalises well to unseen data.

Using PCA, it is less conceptually straight-forward how to apply this method: it is clear that the model complexity we want to determine are the number of principal components, and that PCA must be used somehow to make a prediction of a point that is not in the training set.

We will then introduce a first clear and simple way of performing prediction using PCA that is although not the best one, in order to then explain how to do it properly.

### 5.2.3.1 Row-wise cross-validation

Row-wise cross-validation [19], consists in choosing as test set a point  $x_{(t=i,:)} = x^i$  of dimensions  $< 1.N >$ , and performing PCA on the rest of the data  $X^{(-i)}$ , obtaining an  $N$  by  $N$  matrix called  $PC^{(-i)}$ , whose columns are the PC's. We call  $PC_k^{(-i)}$ , the first  $k$  columns of  $PC^{(-i)}$ . We use the matricial notation of the dimensions  $< a.b >$  to indicate that a matrix has  $a$  rows and  $b$  columns.

We can project the point from the test set, onto the first  $k$  components, through the operation  $(x^i)^t.PC_k^{(-i)}$  and obtain a projected vector of dimensions  $< 1.k >$ . If we re-project back this vector onto the canonical basis, multiplying the last term by  $(PC_k^{(-i)})^t$ , we obtain an estimate of the point in the test set built using only the first  $k$  principal components:  $(x^i)^t.PC_k^{(-i)}.(PC_k^{(-i)})^t$ .

We can then compute, for each order  $k$  of the model, the euclidean distance between the two vectors, the original point and the reconstructed point, which uses only its projection onto the first  $k$  components, for every training set. The predicted residual error sum of squares (PRESS) of order  $k$  is:

$$PRESS_k = \sum_{i=1}^T ||(x^i)^t - (x^i)^t.PC_k^{(-i)}.(PC_k^{(-i)})^t||^2$$

In spite of its simplicity, this approach is not correct because the estimate of  $x^i$ ,  $(x^i)^t.PC_k^{(-i)}.(PC_k^{(-i)})^t$ , is not independent of  $x^i$ . The more PC we use for estimating  $(x^i)$ , the better the reconstruction will be, because we are reconstructing a vector using the projections of that same vector on an orthonormal basis. The predicted residual therefore decreases with the order of the model.

All the other standard cross-validation methods are variants of this simple method, that try to generate estimates of the point that are independent

of the point in question, for example leaving out instead of a point  $x^i$ , one particular dimension  $j$  of the point  $x_j^i$ , and then predict this missing value using PCA. Diana and Tommasi [36] as well as Bro et al [19] make extensive reviews on these somewhat convoluted methods. Bro et al [19] concluded that the method proposed by "Eigenvector" (proposed by a software company), outperformed the other evaluated methods in most circumstances. We will then now present this method, that is also simple enough.

### 5.2.3.2 "Eigenvector" cross-validation

As in row-wise cross-validation, the "Eigenvector" method [19] uses as test set one point  $x^i$ , and iterates across all the dimensions  $j \in [1, N]$  of this vector, to compute the error -the difference between the actual value and the estimate:  $(x_j^i - \hat{x}_j^i)^2$ . The key ingredient of this approach is that the estimate  $\hat{x}_j^i$  is built using the point  $x^i$ , in which the  $j^{th}$  dimension is omitted, noted  $x_{-j}^i$ . This point is a vector of dimension  $< N - 1 >$ . We can project this point onto the  $k \in [1, N]$  dimensional PC basis,  $(PC_k^{(-i)})$ , noted also  $P$ ), by omitting the  $j^{th}$  component of all the PC's computed in the train set, that excludes the point  $X^{(-i)}$ . These truncated  $k$  principal components are noted:  $PC_k^{(-i)}|_{(-j)} = P_{(-j)}$ . The vector of dimensions  $< 1.k >$ , composed by the  $j^{th}$  row of  $PC_k^{(-i)}$  is noted  $p_{(j)}$ .

The predicted residual error is:

$$PRESS_k = \sum_{i=1}^T \sum_{j=1}^N (x_j^i - [(x_{(-j)}^i)^t P_{(-j)} \cdot (P_{(-j)}^t \cdot P_{(-j)})^{-1} \cdot p_{(j)}])^2$$

As we see, despite the complexity of the formula, that the predicted element and its estimate are independent. This "Eigenvector method" can also be written in a more compact form as:

$$PRESS_k = \sum_{i=1}^T \sum_{j=1}^N (x_j^i - [P \cdot [P_{(-j)}]^+ \cdot x_{(-j)}^i]|_j)^2$$

where  $[M]^+$  designates the pseudo inverse of a matrix, and  $v|_j$  designates the  $j^{th}$  component of the vector  $v$ .



## 5.3 Proposed method

### 5.3.1 Motivations

When designing our method, we wanted it to have certain properties that we detail hereinafter.

First, we wanted it to be assumption-free about the distribution of the data, which for the majority of cases is unknown. Certain statistical frameworks like cross-validation and hypothesis testing have this property. However, contrary to cross-validation methods, the hypothesis testing frameworks (like non-parametric statistics), offer an estimate of the significance of the data under a null hypothesis, and this is not the case for cross-validation methods, whose goal is only to minimize the prediction error.

Another important motivation for designing this method, is that when dealing with biological data, we need to be able to make statements about specific time scales, because often this data contains a variety of time-scales, not all of which have the same origin or are necessary of interest. A commonly observed trend is that in biological data, the power decreases like the inverse of the frequency  $1/f$ . Because of this, when one uses Pearson correlation, they are typically dominated by the low frequencies, which can be irrelevant processes: for instance in a typical neuro-physiological recording, one see cells dying and appearing along the recording. We then designed a pre-processing method that is robust to this problem. To this ends, we adapted a method from Schwartz et al. [134].

Before proceeding with the method, we will first introduce separately different steps that intervene in the algorithm.

### 5.3.2 Preprocessing

When we are dealing with variables that are sampled repeatedly from a same object, in order to control for non-stationarities, we center the counts in fixed jitter windows, as we described in the methods section of the second chapter (see figure 2.26).

When we will deal with observations that represent different characteristics of different individuals, we will simply use the mean across the whole population. If there are known subgroups in the populations we can subtract their mean subgroup by subgroup .

### 5.3.3 Shuffling the data: preserve the variance, destroy the covariance

The shuffling we are going to implement consists of permuting randomly the time bins of a given variable inside the jitter window. For point-wise processes, like spike trains, if we randomly reassigned the spikes on a local window around the spike, and then we did some smoothing (like counting or with some Gaussian kernel), the variance would clearly not be preserved.

If we rather permute the counts inside each jitter window (counting the spikes presents in the time bins of length  $t_c$ ), see figure 2.26, this preserves the variance  $V$  of each variable, because the sum is commutative:  $V = \frac{1}{N_b \cdot N_T - 1} \sum_{c=1}^{N_b} \sum_{i=1}^{N_J} (n_i^c - \mu^c)^2$ . However, the shuffling destroys the covariance between the variables.  $N_b$  is the number of jitter windows.

The core idea behind a shuffling test is to see how does one statistic from the data compares with the distribution of statistics from the surrogate data. For example, in this particular case, if we want to test whether the measured correlation between the variables a and b is statistically significant, we compute the distribution of correlations obtained when we shuffle the data  $Q$  times, and see where does the correlation between a and b lies in this distribution (see figure 5.4 C). If it is very far apart from the distribution bulk (p value very small), we can reject the null hypothesis, that states that the correlation between a and b is likely to have been generated by a random fluctuation. The p-value is the proportion of surrogates that have a correlation value higher or equal to the original correlation of these two variables (see figure 5.4).

#### 5.3.3.1 Shuffling the data in a particular subspace

This method consists in doing exactly what we just explained, but considering particular directions of the space of variables. For example, if we project the data onto a subspace generated by two vectors  $\langle PC_x, PC_y \rangle$ :  $X \cdot [PC_x, PC_y]$ , we obtain a matrix of dimensions  $\langle T, 2 \rangle$ , in which each column is a time series of dimensions  $\langle T, 1 \rangle$ . In this method, we assess both whether the variance  $\lambda_x$  is statistically significant with respect the variances of the surrogate data in the direction  $PC_x$ , but also whether the variance  $\lambda_x$  is statistically different from the variance  $\lambda_y$ , by comparing how does the projected variances of the surrogates tend to differ.

For example, if  $\lambda_x = \lambda_y = \frac{\lambda_x + \lambda_y}{2}$ , then a shuffled version of the data in this subspace, that will preserve the total variance  $\lambda_x + \lambda_y$ , will have very similar variances along both subspaces  $\lambda_x = \lambda_y = \frac{\lambda_x + \lambda_y}{2}$ . From this simple scenario, we will deduce two things: first, that the variance of the data  $\lambda_x$  is not statistically different from the variance of the surrogate versions of the data, projected in the direction  $PC_x$ . Second, we will notice that the variance of the component x is neither statistically different from the variance of the component y. This will show us that the data is spherical in these two directions  $PC_x$  and  $PC_y$ . We would arrive at an opposite conclusion if we had for example  $\lambda_x \gg \lambda_y$  (for example figure 5.4 taking  $\lambda_x = \lambda_1$  and  $\lambda_y = \lambda_2$ ).

The question we will now answer is then how to do such shuffling in a given subspace.

After computing the scores in particular directions, we compute their variance  $\lambda_x$  and  $\lambda_y$ . If we project back the scores to the canonical space, right multiplying by  $[PC_x, PC_y]'$ :  $X.[PC_x, PC_y].[PC_x, PC_y]'$ , we obtain a T by N matrix of rank two. We can then apply the shuffle procedure in the canonical basis.

As we know, this shuffle keeps the variance in the canonical basis, but not the total variance in the subspace we are interested. Moreover, shuffling the data this way makes the surrogate data be of rank N again (see figure 5.6 B and C). However, we obtain what we where looking for, which is that the surrogate data is drawn from a spherical distribution.

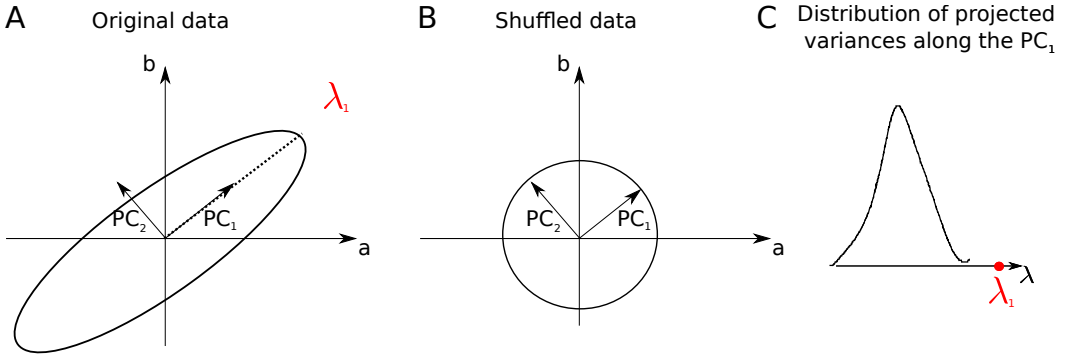
After we shuffle the data ( $\tilde{X}$ ), we project back to the directions of interest  $Sc = \tilde{X}.[PC_x, PC_y]$ . We compute the variances of these scores  $Var(Sc(:, i)), i \in [x, y]$ , and we scale these variances by a constant:  $var(Sc(:, i)) \leftarrow var(Sc(:, i)) \frac{\sum_{i \in [x, y]} \lambda_i}{\sum_{i \in [x, y]} var(Sc(:, i))}$  in such a way that they have the sum of these variances is the same as in the dimensions of interest from the original data.

### 5.3.4 Use of this shuffle to determine the statistical significance of the principal components of the data

We are going to see how to assess whether couples of variables in the original data have correlations that would have been generated by chance if these variables were uncorrelated. We can then quantify where does the actual

value of the correlation between two variables lies in the distribution of correlations obtained by shuffling these two variables, by computing the p-value.

The principal components of the data give us an orthonormal basis in which the directions point to the directions of the highest projected variances. When we shuffle the joint distribution of the two variables  $a$  and  $b$ , we obtain a distribution that is spherical. We can, by projecting again the shuffled distribution on the direction of the first principal component of the original data, determine the projected variance onto the first PC for one shuffle. Repeating the process for  $Q$  shuffles, we can compute the p-value for the first principal component.



*Figure 5.4: Shuffling the data in order to determine the statistical significance of the first principal component. A: Sketch of data having strong correlations. B: Shuffled version of the data in which the variance of the variables is preserved, and the correlation is destroyed. C: Histogram of projected variances along the original  $PC_1$  direction, using different shuffles of the data. In this cartoon, the p-value is very small, and then the first PC is statistically significant.*

Note that we projected the variance on a basis determined in the original data, because the idea is to test whether the particular linear combination of variables we are looking at is statistically significant, in the same way we test whether the correlation between the two variables  $a$  and  $b$  is statistically significant.

One could have also computed the PCA on the surrogate data and de-

termine the eigenvalues, irrespectively of the dimension. In that case we would be testing whether the projected variance is statistically significant with respect to any other joint fluctuation of the measured system, like in Parallel analysis. Remark that in both cases it is essential, in order to make a fair comparison, that the variance of all the variables in the shuffles be equal to the variance of the original data.

### 5.3.5 Signal and noise

The basic assumption when we do such tests, is that the PC's that have a variance that is significantly different of what could be expected by chance are called signal, and the others are called noise. According to Bro and al. [19] "the noise can be loosely defined as any specific variation in a measurement which is not correlated with any other variation in the measured data".

We place ourselves in the case in which we test the statistical significance of one direction, i.e., we project the shuffled data onto the same PC determined in the original data set. Now, imagine that in the figure 5.4, the first PC corresponds to signal and that the second PC is noise. When we compare the values of the projected variances onto the  $PC_1$  and the  $PC_2$  for the original data and for the shuffled data (see A and B, figure 5.4), we see that the projected variance of the  $PC_1$  in the shuffled data is much smaller than in the original data, and that the projected variance onto  $PC_2$  is bigger in the case of the shuffled data with respect to the original data: when shuffling the data, there is a transfer of variance from the signal, -in this case the  $PC_1$ - to the noise -the  $PC_2$ -.

We then understand that we must adopt an iterative approach to evaluate the statistical significance of the PC's: at first iteration, we observe whether the eigenvalues of the original data fall, one by one, inside the respective histograms of projected variances. If that is indeed the case, it means that there are no correlations in the data which are statistically significant. However, if it is not the case, the existence of this transfer of variance implies that we must be careful when evaluating the statistical significance of subsequent PCs. More precisely, in order to establish the statistical significance of the second PC's with respect to the next PC's, one must first **subtract** out from the data the contribution of the first PC, and then shuffle the data (and then rescaling by a constant), to see if it is spherical on those PC's. In this

chapter, when we use the term subtract a subspace from a data set, we mean project onto its orthogonal complement. This approach was introduced by Schwartz et al. [134].

For example in the following figure 5.5 we see that there are three dimensions. After subtracting the first principal component to the data (in green in A and B). After shuffling the data on the subspace formed by  $\langle PC_2, PC_3 \rangle$ , we see (in B) that the projected variances are undistinguishable from the projected variances that would have been generated by two uncorrelated variables  $PC_2$  and  $PC_3$  (in blue).

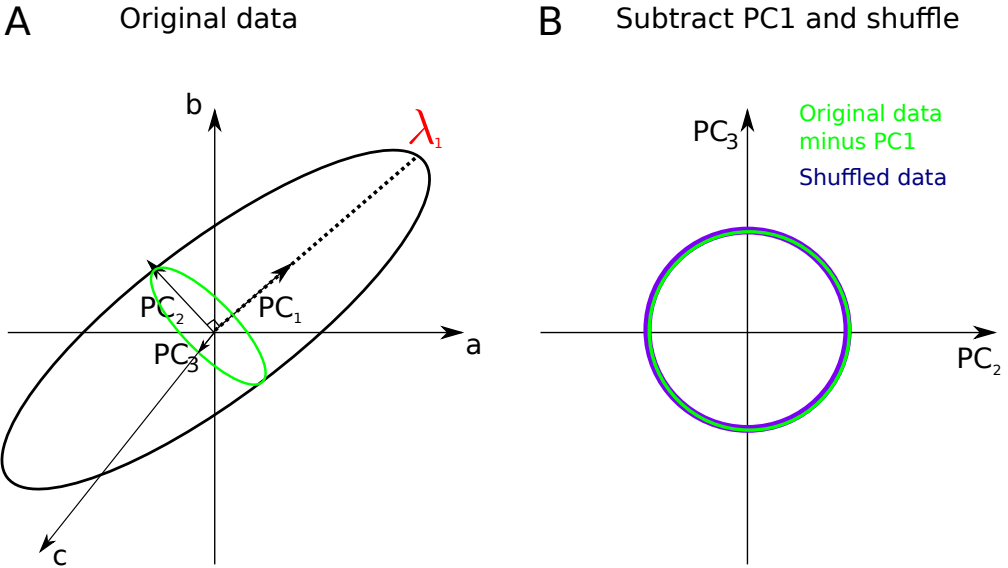


Figure 5.5: Subtracting  $PC_1$  to the data and then shuffling the data in the  $\langle PC_2, PC_3 \rangle$  subspace shows that  $\lambda_2$  is not statistically significant with respect to  $\lambda_3$ . In A, original data. The principal components find the directions of more variance. Given that the  $PC_1$  is statistically significant, we subtract its projection from the data (green circle). In B, zoom and rotated version of data having excluded the contribution from the first PC (green). In blue, shuffled version of the data. As we see in this sketch, the variances projected are very similar. This means that the second and the third PC might be considered as noise.

A fact that is often neglected in the literature, is that the variance can be significantly higher than the one that we would obtain by chance (amplified variance), but also significantly smaller (compressive variance) at the end of

the eigenspectrum, as pointed out by Schwartz and al [134]. In our iterative procedure, we will then eliminate -if necessary- PC's from both extremities of the eigenspectrum. The iterative procedure stops when, for each of the remaining PCs, the eigenvalue of the data is inside a confidence interval of the surrogate distribution:  $[\alpha; 100 - \alpha]\%$ .  $\alpha$  is the false alarm rate (see figure 5.7 D).

### 5.3.6 Detailed procedure

Lets call  $X$  the data after having binned the spikes, centered the counts in fixed windows (as described previously), and z-scored (i.e. divided each column representing the activity of one neuron, by its standard deviation).  $X$  has dimensions  $N_T$  by  $N$ . This method is an iterative procedure, we will first simply describe the algorithm and then comment it.

**inputs:**

$X, t_c, N_J, Q, \text{DimProjOut}$ .

$Q$  is the number of shuffles.

$\text{DimProjOut}$  is a list of numbers that encompasses all the PC we want to subtract of from the data.

Example:  $[1, 2, 3, N]$ . In the first iteration, we obviously let the list empty:  $[]$ .

**outputs:**

$\lambda^o, \{\lambda_i^s\}_{i=1, \dots, N}^{s=1, \dots, Q}$

$\lambda^o$  is the eigenspectrum of the original data.

$\{\lambda_i^s\}_{i=1, \dots, N}^{s=1, \dots, Q}$  are the eigenspectra for every shuffle.

**algorithm:**

```

 $[PC^o, \lambda^o] = \text{PCA}(X)$ 
 $Sc = X.PC^o$ 
 $DimProj = \{1, \dots, N\} \setminus DimProjOut$ 
 $X^* = Sc(:, DimProj).PC^o(:, DimProj)^t$ 
 $\forall s = 1, \dots, Q$ 
     $\forall n = 1, \dots, N$ 
         $X^{shuff}(:, n) = \text{Permute}_{(N_J)}(X^*(:, n))$ 
         $Sc^{shuff} = X^{shuff}.PC^o$ 
        if  $n \in DimProj$ 
             $\lambda_n^s = \text{var}(Sc^{shuff}(:, n)) \frac{\sum_{i \in DimProj} \lambda_i^o}{\sum_{k \in DimProj} \text{var}(Sc^{shuff}(:, k))}$ 
        else :  $\emptyset$ 
Compute p-value of  $\lambda_{i=DimProj(1)}^o$  or  $DimProj(end)$ 

```

**comments:**

$\text{Permute}_{(N_J)}(v)$  reassigns randomly the time counts of vector  $v$  inside the blocks of length  $N_J$ .

Dimensionality:

$Sc$ : scores. Dimension:  $< N_T.N >$ ,  $Sc^{shuff}$ :  $< N_T.N >$ .

$\text{rank}(X^*) = \text{card}(DimProj)$  Ex:  $DimProjOut = \{1\}$ , and  $DimProj = \{2, 3, \dots, N\}$ , then  $\text{rank}(X^*) = N - 1$ .

### 5.3.7 Explanation

If the data is spheric, there is not a direction that has more variance than the others, so that when we shuffle the data, the projected variance of the shuffled data in the original PC basis is not significantly different from the original variance. Now, lets suppose that we found that the first PC was significant (as in figure 5.4), and that the remaining eigenspectrum didn't fell into the tube, which would mean that there is nothing else to explain. When we talk about the tube we mean the tube comprising  $[\alpha\%, 100 - \alpha\%]$  of the variances of the surrogate data (see 5.7). In the second iteration, we are going to subtract to the data the first PC. If one does not remove the first PC, there is a transfer of variance from the signal to the noise, as



shown in 5.4 (look at the change of variance of the second PC before and after shuffling).

To test whether the second PC is statistically significant, one has to remove the first PC from the data and then test the specificity of that modified data. This means that we have to shuffle the data in the subspace spanned by  $\langle PC_2, PC_3, \dots, PC_N \rangle$ . To do so in a way that preserves the remaining variance of each cell, we perform the shuffle on  $X^*$ , which is the data in the canonical basis to which we subtracted the first PC.

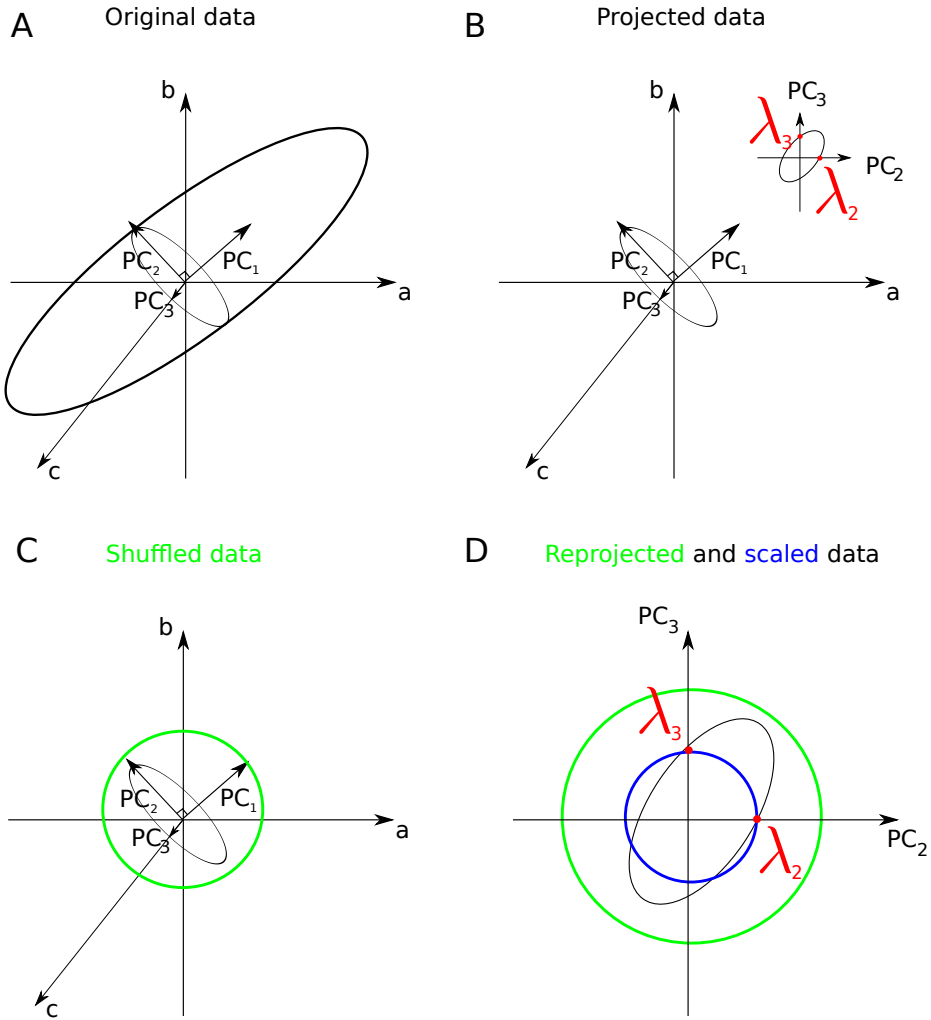


Figure 5.6: Method summary. A: Schematic of data in 3D, and associated PCs. B: subtract the projection of  $PC_1$  from the data. Insight projection of the data on the  $\{PC_2, PC_3\}$  basis,  $\lambda_2, \lambda_3$ : associated eigenvalues. C: Shuffled data: When we shuffle the data, it becomes full rank again. D: We re-project back the data onto the PCs and rescale the eigenspectrum by a constant in such a way that the projected variances have the same value as in B. In this case the second PC doesn't have significantly more variance than the  $PC_3$ .

However, as we perform the shuffle on  $X^*$  two things happen: on one side, the data is not confined anymore to the subspace  $\langle PC_2, PC_3, \dots, PC_N \rangle$  and the variance projected on the axis  $\{PC_2, PC_3, \dots, PC_N\}$  scales multiplicatively. To overcome these two drawbacks we respectively first reproject the shuffled data in the directions  $PC_2, PC_3, \dots, PC_N$  and then we rescale the variance in each of those dimensions by a factor  $\frac{\sum_{i \in DimProj} \lambda_i^o}{\sum_{k \in DimProj} var(Sc^{shuff}(:,k))}$ .

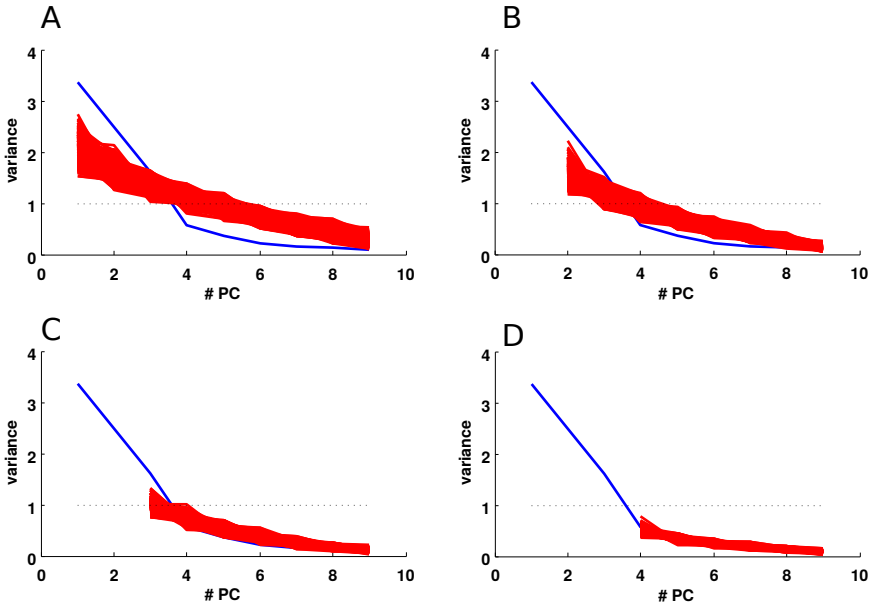


Figure 5.7: Determining the significant components consists in taking out iteratively the components that have significant variance, and making the eigenspectrum of the remaining components to lie into the "tube". A: Nil iteration. B: First iteration. C: Second iteration. D: Third iteration. In blue: data eigenspectrum. In red: projected variances of the shuffled data. Number of shuffles: 1000. Artificially generated data, normal multivariate of dimension 9 and of covariance matrix called (1) in [114] (see in a subsequent section).  $T$ : 27 observations. From this analysis we conclude that there are 3 significant principal components.

## 5.4 Comparing different methods

### 5.4.1 Comparing methods with different data-bases

A standard procedure in the literature is to evaluate the methods in standard databases. These data bases are described in the table 5.1, and are publicly available [48],[86]. Peres-Neto et al.[114] do a very thorough comparison with known biological databases. However, we didn't used the same data sets they used, because not all of them are publicly available, and from the ones we could have access to, they have missing data and a small number of observations.

We didn't applied local preprocessing to this data, because all these databases include conditionally independent samples. We only z-scored the data. The kind of preprocessing we described in figure 2.26 is more adapted to multivariate continuous process, or to several samples from the same entities.

Data set	Variables	Observations
BUPA	7	345
COW	10	50
IRIS	4	150
WINE	13	178

Table 5.1: Description of different databases.

We then applied to all of these data sets, many of the methods described previously.

Method   Data set	BUPA	COW	IRIS	WINE
MAP	1	1	1	3
Parallel	1	3	1	3
Eigenvect	1	1	2	4
Our method	1	2	3	3

Table 5.2: Different number for different data sets, using different methods.

We also tried using factor analysis, and the cumulative modes of the correlation matrix, but these techniques were not conclusive. In both cases,

in factor analysis and in cross-validation of the correlation modes, in spite of evaluating the model on a test set, the more factors we added, and the higher the log-likelihood was (or respectively the fraction of explained variance). One could always put an arbitrary cut-off to the fraction of explained variance -for example- and say, that when it attains 95 %, this defines the number of components. In my opinion, this is not a very principled procedure: on one side this means that the cross-validation step doesn't work, and that we are reintroducing a rule-of-thumb. On another side, given that the cross-validation doesn't fullfills its function, we could directly use other more direct and ad-hoc rules like the Bayesian information criterion (BIC) to determine the complexity of the model, but we didn't do it, since is not very satisfactory.

As Ferré [41] highlights, comparing different methods, in absence of a ground truth, is a current procedure, but it is a bit worthless, given that we don't have a ground truth, because all the methods might be inaccurate. In a subsequent paragraph, we will numerically generate data from a known distribution, and we will observe how does these methods perform on it.

Even more, one might think that even the notion of ground truth is not totally transparent. In the case of the IRIS data set, it is known that there are three types of IRIS flowers: Setosa, Versicolour, and Virginica. Two of these species (Versicolour and Virginica) are very similar. When we do PC with our method it appears that even if the second and the third eigenvalues are not in the envelope of projected variances, they are not far from being on it. It is then patent that even if our method can pick those the differences between the three species, it could have been that even if those differences existed, they might have been so small that there are statistically undistinguishable. In fact, we have the same problem when we generate artificial databases, because we can not know in advance how a qualitative difference in the variance of two PC's will translate into a categorical statement of significant or non statistically significant.

Conversely, even if it is good to know whether one PC is statistically significant, we have to remember that on the first place, a PC can be significant and not be interesting, i.e. to not explain much more shared variance than what is explained by chance.

5.4.2 Comparing methods with different artificially generated data sets

5.4.2.1 Generating ground-truth data sets

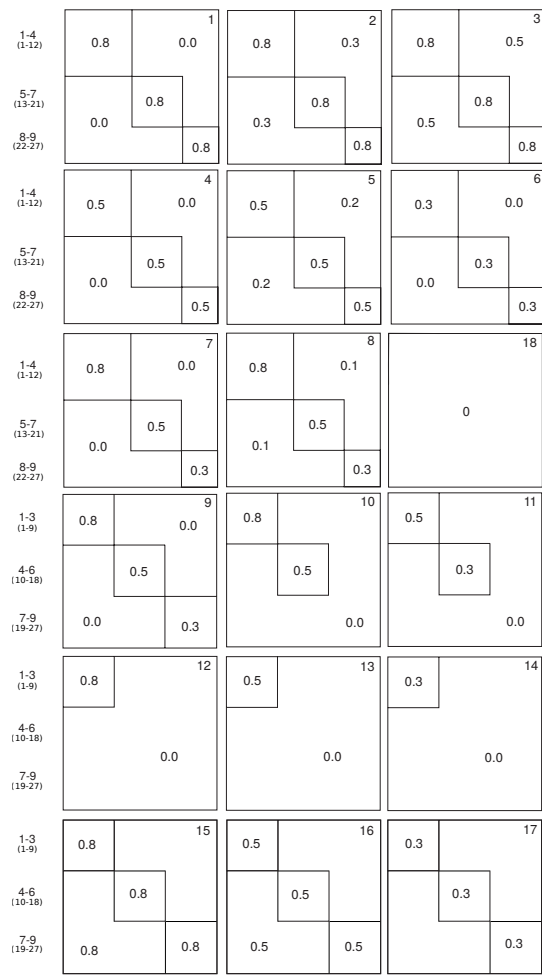


Figure 5.8: Correlation matrices used to generate data. The diagonal elements are omitted. The matrices can have two sides: 9 by 9 or 27 by 27. At left, the variable numbers. The numbers inside the matrices are correlations. Adapted from [114].

Following the approach proposed by Peres Neto et al [114], we generated artificial data from a known distribution. In our case, we limited ourselves to generate data from normal multivariate correlated variables (Peres Neto et al also use exponential and exponential cube). Using the correlation matrices, proposed by them, -see figure 5.8-, we generated correlation matrices of size  $N = 9$  or  $N = 27$ , and we also varied the amount of observations to respectively  $3N$ ,  $6N$  and  $9N$  for each matrix.

After having generated the 18 correlation matrices  $C$ , we rotate them:  $C^* = PCP'$ . We use the same unitary rotation matrix  $P$  for every matrix of the same size. We obtain the rotation matrices, by computing the eigenvectors of a symmetric matrix:  $M = N+N'$ , where  $N_{ij} \sim N(0, 1), i, j \in \{1, N\}$ . We eliminated the matrices 2,3,5 and 8, because it is not totally clear what is their ground truth (1 or 3 PCs ?).

5.4.2.2 Number of PC for different methods

In order to evaluate the performance, we counted as a hit if the rules gave exactly the putative good answer, that we call "ground truth". We call "*corr #*" the number of the correlation matrix shown in the figure 5.8.

rule   corr #	1	4	6	7	9	10	11	12	13	14	15	16	17	18	Performance
MAP	3	0	0	1	2	0	0	1	0	0	1	1	1	0	<b>42,8 %</b>
Parallel	3	2	4	3	3	4	4	3	5	4	1	1	1	5	<b>42,8 %</b>
Eigenvector	3	3	2	2	2	1	1	1	1	1	1	1	1	1	<b>57,1 %</b>
Our Method	3	2	0	1	2	0	0	1	0	0	1	1	1	0	<b>42,8 %</b>
Ground truth	3	3	3	3	3	2	2	1	1	1	1	1	1	0	100 %

Table 5.3: Different number of statistically significant components for different data sets, using different rules.  $N: 9$ ,  $T: 27$  (number of observations)

rule   corr #	1	4	6	7	9	10	11	12	13	14	15	16	17	18	Performance
MAP	2	0	0	1	0	0	0	1	0	0	1	1	1	0	<b>35,7 %</b>
Parallel	3	3	2	3	4	3	3	2	2	4	1	1	1	4	<b>42,8 %</b>
Eigenvector	3	3	1	3	2	2	1	1	1	1	1	1	1	1	<b>71,4 %</b>
Our Method	3	1	1	3	2	3	0	2	1	0	1	1	1	0	<b>50 %</b>
Ground truth	3	3	3	3	3	2	2	1	1	1	1	1	1	0	100 %

Table 5.4: Different number of statistically significant components for different data sets, using different rules.  $N: 9$ ,  $T: 54$  (number of observations)

## 5 Determination of the number of statistically significant principal components

rule   corr #	1	4	6	7	9	10	11	12	13	14	15	16	17	18	Performance
MAP	3	2	0	0	0	2	0	0	0	0	1	1	1	0	<b>42,8</b> %
Parallel	3	3	3	3	4	2	4	3	4	5	1	1	1	7	<b>57,1</b> %
Eigenvector	3	2	1	3	2	3	1	1	1	1	1	1	1	1	<b>57,1</b> %
Our Method	3	3	1	4	4	2	1	2	1	0	1	1	1	0	<b>57,1</b> %
Ground truth	3	3	3	3	3	2	2	1	1	1	1	1	1	0	100 %

Table 5.5: Different number of statistically significant components for different data sets, using different rules.  $N: 9$ ,  $T: 81$  (number of observations)

rule   corr #	1	4	6	7	9	10	11	12	13	14	15	16	17	18	Performance
MAP	3	4	3	4	3	2	2	2	1	0	1	1	2	0	<b>64,2</b> %
Parallel	3	3	3	3	3	5	5	5	1	8	1	1	1	11	<b>64,2</b> %
Eigenvector	3	3	3	4	3	2	2	1	1	1	1	1	1	1	<b>85,7</b> %
Our Method	3	3	3	3	3	2	2	3	1	1	1	1	1	0	<b>92,8</b> %
Ground truth	3	3	3	3	3	2	2	1	1	1	1	1	1	0	100 %

Table 5.6: Different number of statistically significant components for different data sets, using different rules.  $N: 27$ ,  $T: 81$  (number of observations)

rule   corr #	1	4	6	7	9	10	11	12	13	14	15	16	17	18	Performance
MAP	3	3	3	3	3	2	2	1	1	1	1	1	1	0	<b>100</b> %
Parallel	3	3	3	3	3	5	3	6	4	3	1	1	1	12	<b>57,1</b> %
Eigenvector	3	3	3	3	3	2	2	1	1	1	1	1	1	1	<b>92,8</b> %
Our Method	3	3	3	3	3	6	3	7	1	1	1	1	1	0	<b>78,5</b> %
Ground truth	3	3	3	3	3	2	2	1	1	1	1	1	1	0	100 %

Table 5.7: Different number of statistically significant components for different data sets, using different rules.  $N: 27$ ,  $T: 162$  (number of observations)

rule   corr #	1	4	6	7	9	10	11	12	13	14	15	16	17	18	Performance
MAP	3	3	3	3	3	2	2	1	1	1	1	1	1	0	<b>100</b> %
Parallel	3	3	3	3	3	6	5	13	6	2	1	1	1	13	<b>57,1</b> %
Eigenvector	3	3	3	3	3	2	2	1	1	1	1	1	1	1	<b>92,8</b> %
Our Method	3	3	3	3	3	12	3	1	2	1	1	1	1	0	<b>78,5</b> %
Ground truth	3	3	3	3	3	2	2	1	1	1	1	1	1	0	100 %

Table 5.8: Different number of statistically significant components for different data sets, using different rules.  $N: 27$ ,  $T: 243$  (number of observations)



## 5.5 Determining the different number of statistically significant components in different experiments during desynchronised state

We then used these methods in order to compare how many significant principal components where there in the recordings of spontaneous activity. We applied to the data the usual preprocessing (local centering), with the usual parameters  $t_c = 0.1s$ , and  $N_J = 10$ . Given the time count chosen there where between 3000 and 6000 observations. Because of the computational complexity of our method and of the "Eigenvector" method, we limited the number of observations for this methods to 1000 observations for all data sets.

Data set	N neurons	MAP	Parallel	Eigenvector	Our method
Set 1	100	2	20	1	1
Set 2	116	3	22	1	3
c065a	65	1	11	1	2
c065b	47	1	6	1	6
c043	67	2	19	1	0
c048	44	1	12	1	0
c008b	93	2	32	1	0

*Table 5.9: Different number of statistically significant components for different data sets, using different rules.*

## 5.6 Discussion and conclusions

### 5.6.1 Results from the comparison with simulated data

As we see, from the previous results, our rule to determine has comparable results to the other state of the art rules, but it doesn't outperforms them. The rule that seems to have better results across conditions is the eigenvector rule. Our method only does better in one of the six tests. When we compare

our method with parallel analysis, our method always performs similar or better than parallel analysis. In general all these methods do better when we increase  $N$  the size of the matrix and also the number of available samples per matrix. The problem of the "Eigenvector" method is that it is very time consuming, so it becomes intractable for long enough data sets. The method called MAP, in its revised version [153] shows excellent performance when there is enough data, but has the worst performance with small amounts of data.

If we observe the errors, we can see that for small amounts of data, our method underestimates the number of PC, and for big amounts of data, our method overestimates the number of PC.

### 5.6.2 Results from the analysis of spiking data from desynchronised state

The parallel analysis gives a number of factors that is not coherent with the result of the three other methods. The first reason that might explain this result is that this analysis is simply based on the fact that the data eigenvalues are above or below the mean eigenvalues of the shuffled data. In the original version of parallel analysis [69], there is no notion of distribution width.

When the eigenvalues of the are very above from the mean eigenvalues of the shuffled data, there is no doubt about the significance. However, when the decay of both "eigensprectra" is parallel and has a small distance between the two, then it is very unreliable.

The three methods, MAP, "Eigenvector" and our method, display results that are not equal but that have a small number of statistically significant components.

### 5.6.3 Advantage of the preprocessing to evaluate non-stationarity of the data

The preprocessing introduced in the second chapter (see 2.4.2) and also mentioned in this chapter is very interesting because it allows to evaluate the correlations at different time scales. We choose to compare our method to simulated data set, following [114], because it provides a way of assessing some of the dimensions that these methods have to deal with: data size,

number of variables, number of significant PCs. However, one of the limitations of this evaluation framework, is that it only generates stationary data. One of the possible future directions would be to test this method and this preprocessing with controlled non-stationary data, to see how well it captures the correlations. Indubitably, when comparing our method with the other methods, we should also apply the preprocessing to the data before feeding it to the other methods.



# Chapter 6

## General conclusion and discussion

HIGHLIGHTS

- We recapitulate the main results of this work.
- We open the discussion on the different research possible future paths.

### Contents

<b>6.1</b>	<b>Main subject of this work</b>	<b>246</b>
<b>6.2</b>	<b>Results</b>	<b>247</b>
6.2.1	Competitive activity during desynchronised state	247
6.2.2	Modelling of competitive activity	247
6.2.3	Probing the relationship between local and long range connectivity	248
6.2.4	Methodological contributions	248
<b>6.3</b>	<b>Discussion</b>	<b>249</b>
6.3.1	A new form of amplification: feed-forward amplification	249
6.3.2	Overall view on amplification	251

## 6.1 Main subject of this work

This thesis could also have been called "connectivity and dynamics in cortical circuits". Our efforts have been targeted to understand the link between the observed activity and circuit mechanisms by which this activity could have been generated.

One of the angles by which we tackled this problem both for the data analysis as well as for the modelling, was to use the framework of Ganguli et al [46] and Murphy and Miller[105], in which there is a simple interpretation of how does connectivity shapes spontaneous activity . In this framework, it is considered that local circuits receives unstructured activity from the outside, and that the connectivity of these circuit constrains the range of the expressed activity patterns. Therefore, this motivates the use of multivariate analysis tools like PCA, to observe whether there are specific patterns of activity that are specially amplified.

Using PCA, we discovered the existence of population-wide temporal unidimensional structure during the asynchronous state, that we characterised. Seeing that there is one pattern of activity that has more variance than the others, the formalism of the Schur decomposition gives us a first hint of one characteristic of the effective connectivity matrix. Using a previous model of competition [155], we proposed two models which could reproduce such competitive dynamics. One of the two models, TCA, has a particular asymmetry in the connectivity, and generates dynamical predictions that match the data. Our interpretation of the most important prediction of the model, the lag in the ccg between the two populations, is that it is a signature of non-normal amplification.

We then made a second study in which we also explored a direct link between the connectivity and the amplification. We concentrated on modelling and simulating how does the positive correlations between neurons increase as the fraction of shared input that these cells receive increase. This modelling gives support to the anatomical evidence brought by N. Morgenstern showing that in layer four and between layer four and layer 2/3, specific sub-networks, with similar receptive fields are thought to be preferentially connected.

## 6.2 Results

### 6.2.1 Competitive activity during desynchronised state

After comparing what previous studies had shown about the temporal structure of spontaneous activity - mainly during the more synchronised states-, we identified that during the desynchronised states there are coherent uni-dimensional fluctuations such that the population firing rate stays constant and alternatively half of the population increases its firing rate while the other half decreases its firing rate.

We characterised the competitive activity: we first saw that the direction of the competition remains still over time in all the recordings. When studying the time scale of the competition, we found that there is competition at many time scales. Also, we understood that each group can be labelled independently given the firing properties of the cells: in one group, the neurons fire more tonically while in the other group they fire in a more bursty way. Finally, we identified that the competition happens locally: neurons with positive and negative loadings are intermingled in a single recording shank.

### 6.2.2 Modelling of competitive activity

Given that the randomly connected network [123], which is the standard model of desynchronised state, doesn't show competitive amplification, we proceeded to model the competitive activity, trying to find a circuit motif that reproduced the dynamical observables of the data. We ended finding two possible circuits based on two different mechanisms: NCA -normal competitive amplification- and TCA -transient competitive amplification- based on non-normal amplification. TCA reproduces better many dynamical observations of the data and in particular the asymmetry in the cross-correlogram between the two populations  $E_1$  and  $E_2$ . In a second moment, we asked how could we generate such low dimensional dynamics in a high dimensional network. We then extended the low dimensional TCA motif to high dimensions. In high dimensions there are as many possible amplified directions as there are dimensions, but the directions that dominate the dynamics are still the dimensions of the low dimensional model. This model confirms the previous predictions of the low dimensional model, and adds

high dimensional dynamical predictions that also reproduce well the data (graded correlation matrix, population raster, PC1 loadings, histogram of pairwise correlations).

### 6.2.3 Probing the relationship between local and long range connectivity

In a separate anatomical study, done in collaboration with N. Morgenstern and L. Petreanu, we studied the spatial specificity of long range thalamocortical projections. N. Morgenstern and L. Petreanu proposed a technique to quantify the amount of shared axons on pairs of neurons, using patch clamping and optogenetic stimulation. Stimulating optogenetically long range axons with minimal intensity, one can observe how pairs of cells covariate and then know if they receive shared inputs or not, and separately measure whether those cells are connected or not.

Using this technique, they showed spatial selectivity of the thalamocortical projection into layer four: connected pairs of cells tend to receive more shared inputs from axons coming from the visual thalamus. I modelled the experiment in two different ways: a more analytically grounded, and a more heuristic one, based solely on simulations, which made different assumptions on the way the experiment was run. The retained approach, using simulations, shows how a difference in the mean number of shared axons leads to differences in the way the number of correlated sites grow with the number of recruited axons, and this simulation resembles qualitatively to what we observe in the data.

### 6.2.4 Methodological contributions

#### 6.2.4.1 Data analysis

On the data analysis side, we came out with some methodological approaches that are useful when dealing with high dimensional data. The three most relevant are respectively:

- The pre-processing of the data consisting in centring the binned spike trains in fixed jitter windows. This simple procedure controls for non-stationarities, puts an upper-bound on the time scale of the correlations and cuts the data in a very useful way for cross-validating models on the



data.

- A statistical framework to assess the sufficiency and necessity of particular subspaces of activity in order to explain the correlation matrix.
- A new method to evaluate the number of relevant principal components in a data set.

### 6.2.4.2 Modelling of recurrent networks

On the modelling side, we propose a way of modelling high dimensional linear recurrent networks, which is fairly general: make a low dimensional circuit with suitable dynamics and expand it to high dimension, and then add noise on the connectivity in a principled way. Doing so, the low dimensional dynamics are carried from the low dimension into the high dimensional representation, but in a way that the high dimensional model is full rank and shows heterogeneity in the connectivity.

The condition we impose on the noise makes the Schur decomposition of the total matrix (deterministic plus noise connectivity matrix), be the sum of the Schur decomposition of both connectivity matrices. This condition gives a simple and clear dynamical interpretation of the high dimensional network.

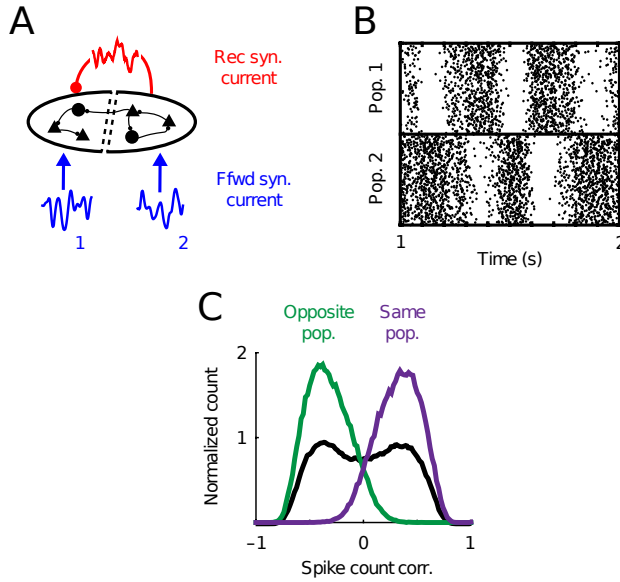
## 6.3 Discussion

### 6.3.1 A new form of amplification: feed-forward amplification

We showed in this work that adding structure to the connectivity, we could reproduce the competitive nature of the desynchronised state. However, other modelling choices are also possible. Another possibility to generate competitive could have been that different populations received anti-correlated input, which is totally plausible, but trivial to model. A less trivial alternative was recently proposed by Rosembaum et al [127]. Starting from the randomly connected balanced network, they plugged two uncorrelated non overlapping feed-forward sources of white noise on two arbitrary halves of the population (figure 6.1 A).

In this case, they observed that in each half of the population the balance is broken, and the activity in each of the halves is synchronous (figure 6.1

B). Interestingly, they observe that the activity between the two halves is negatively correlated [6.1](#), in such a way that even if locally -at the level of a single halve of the network- the balance is broken, at the network level, the asynchronous state is preserved, because the mean correlation is close to zero (figure [6.1 C](#)).



*Figure 6.1: Two independent selective feed-forward inputs cause a break in the asynchronous state. A: two independent feed-forward inputs project to two non overlapping halves of the population. B: Competitive activity appears in this network. C: histogram of pairwise correlations of each half of the population and of the all population. This figure was adapted from Rosenbaum et al. [\[127\]](#).*

Rosenbaum et al [\[127\]](#) argue that this break in the balanced state is provoked by the fact that a same population can not track both inputs at the same time, so that the shared current inherits shared fluctuations from their feed-forward inputs in each population, and this somehow introduces a competitive dynamic between the two populations.

Although this study was published only a few weeks ago, so that we did not have time to properly consider its implications for our work, it is clearly relevant to our investigation of different mechanisms capable of generating amplification in recurrent networks. Our current view is that the mechanism

proposed in [127] is qualitatively distinct from the ones we have considered. Using the framework developed in [123], it can be understood that the kind of amplification presented in [127] results from a specific interaction between the feed-forward and the recurrent connectivity in the network. We could thus call it 'feed-forward amplification'. It is distinct from the models examined in this thesis, in which we have focussed on the effect of the recurrent connectivity on amplification, and have therefore assumed that the external input was independent for all neurons. Qualitatively, feedforward amplification arises when the feed-forward connectivity tries to generate patterns of activity in the recurrent network that are not favoured by the recurrent connections, i.e., when the feed-forward input is close to orthogonal to the range of the recurrent connectivity matrix. In the balanced regime, the cancellation between excitation and inhibition implies that overall recurrent input needs to cancel the overall external input. When the previous relationship between the feedforward and the recurrent connectivity holds, this ongoing cancelation requires very large fluctuations in certain patterns of network activity, i.e., amplification. From a quantitative perspective, however, feed-forward amplification is equivalent to an increase in variance in the external inputs onto particular patterns of recurrent network activity. How this external noise is processed by the recurrent network is still determined by the two kinds of mechanisms discussed in our thesis.

### 6.3.2 Overall view on amplification

In the modelling chapter we discussed ways by which the cortical circuits could amplify the thalamic inputs, mainly based on the connectivity pattern of the recurrent circuit.

The first one, initially proposed by Douglas et al [38] in 1995, is normal amplification, which consists in decreasing the negative feed-back of a network towards the instability. The most characteristic signature of non-normal dynamics is the trade-off between amplification and speed.

The second kind of amplification, non-normal amplification, was proposed by Murphy and Miller [105] in 2009, and consists in trading positive feedback against feed-forward gain in the Schur domain.

These two kinds of amplification are not easy to disentangle, because to do so we must observe the data, the amplified patterns, and other dynamical features, and try to make models using normal and non-normal amplification

to see if there is some robust feature of the data that is idiosyncratic of one of these mechanisms.

In the continuation of the work by Murphy and Miller [105] and Hennequin [64], we showed how to build high dimensional networks that amplify certain patterns of activity, through the recurrent connectivity, and to accommodate the model predictions to the variability of the different neurons.

In the two studied models (NCA and TCA), we didn't considered the case of shared input. The inputs to all of the three populations were supposed to be independent and of the same variance. The feed-forward amplification [127], offers a third possible mechanism by which amplification might be happening in the cortex. Besides, these three mechanisms are not mutually exclusive. To prove it, we must first recall, quoting Douglas et al [38], that "connections arising from the LGN make up less than 10 % of the excitatory synapses formed with the recurrent input neurons of layer IV". Second, we also must cite the seminal study by Lien and Scanziani [87], showing how the recurrent amplification in the cortex amplifies multiplicatively the signals coming from the thalamus onto layer IV (see figure 4.2).

In the chapter of anatomy, we discussed recent evidence pointing at the fact that shared thalamo-cortical axons tend to connect more with pairs of cells that are connected. In the light of the recent study [127] on feedforward amplification, it is interesting to consider the effect that this pattern of connectivity would have on the potential amplification of thalamic signals by the visual cortex. It is currently difficult to predict whether the data in our study [104] would lead to amplification of the kind described in [127]. If our current understanding is correct, this would be the case if the degree of segregation of thalamic input to different cortical subnetworks is much larger than the corresponding segregation of these groups by local recurrent cortico-cortical connections. However, if the degree of segregation is similar, this would not lead to feedforward amplification. Given that the current data only shows that the segregation of thalamic-cortical and cortico-cortical connections goes in the same direction (but not the relative magnitude of the two types of segregation), the question is still open.

To finish we will recall on the importance of the study of amplification of spontaneous activity both during anaesthetised and in behaving animals, to the extent that it might reveal the constraints that connectivity impose on the dynamics, and therefore it might give us hints on the computations that neural circuits are performing.

# Bibliography

- [1] A Agmon and B W Connors. Correlation between intrinsic firing patterns and thalamocortical synaptic responses of neurons in mouse barrel cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 12(1):319–329, 1992. (Cited on page [72](#).)
- [2] Yashar Ahmadian, Francesco Fumarola, and Kd Miller. Properties of networks with partially structured and partially random connectivity. *arXiv preprint arXiv:1311.4672*, pages 1–34, 2013. (Cited on page [127](#).)
- [3] Asohan Amarasingham, Matthew T Harrison, Nicholas G Hatsopoulos, and Stuart Geman. Conditional modeling and the jitter method of spike resampling. *Journal of neurophysiology*, 107(2):517–31, 2012. (Cited on page [81](#).)
- [4] D J Amit and N Brunel. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)*, 7(3):237–252, 1997. (Cited on page [27](#).)
- [5] D. Attwell and S. B. Laughlin. An energy budget for signaling in the grey matter of the brain. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 21(10):1133–45, 2001. (Cited on page [19](#).)
- [6] E. H. Baeg, Y. B. Kim, K. Huh, I. Mook-Jung, H. T. Kim, and M. W. Jung. Dynamics of population code for working memory in the prefrontal cortex. *Neuron*, 40(1):177–188, 2003. (Cited on page [100](#).)
- [7] W. Bair, E. Zohary, and W. T. Newsome. Correlated firing in macaque visual area MT: time scales and relationship to behavior. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(5):1676–97, 2001. (Cited on page [32](#).)
- [8] Peter Bartho, Hajime Hirase, Lenaic Monconduit, Michael Zugaro, Kenneth D. Harris, and Gyorgy Buzsaki. Characterization of Neocortical Principal Cells and Interneurons by Network Interactions and Extracellular Features. *Journal of Neurophysiology*, 92(1):600–608, 2004. (Cited on page [142](#).)
- [9] M.S Bartlett. Tests of significance in factor analysis. *British Journal of Statistical Psychology*, 3(2), 1950. (Cited on page [217](#).)

## Bibliography

---

- [10] Francesco P Battaglia, Gary R Sutherland, and Bruce L McNaughton. Hippocampal sharp wave bursts coincide with neocortical "up-state" transitions. *Learning & memory (Cold Spring Harbor, N.Y.)*, 11(6):697–704, 2004. (Cited on page 20.)
- [11] R Ben-Yishai, R L Bar-Or, and H Sompolinsky. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–3848, 1995. (Cited on pages 100 and 135.)
- [12] Corbett Bennett, Sergio Arroyo, and Shaul Hestrin. Subthreshold mechanisms underlying state-dependent modulation of visual responses. *Neuron*, 80(2):350–357, 2013. (Cited on pages 14, 15, and 20.)
- [13] Orban G Lengyel M Fiser J Berkes P. Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment. *Science (New York, N.Y.)*, 331(January):83–88, 2011. (Cited on page 33.)
- [14] Edgar J Bermudez-Contreras, Andrea Gomez-Palacio Schjetnan, Arif Muhammad, Peter Bartho, Bruce L McNaughton, Bryan Kolb, Aaron J Gruber, and Artur Luczak. Formation and Reverberation of Sequential Neural Activity Patterns Evoked by Sensory Stimulation Are Enhanced during Cortical Desynchronization. *Neuron*, 79(3):555–566, 2013. (Cited on pages 39 and 74.)
- [15] Christopher M. Bishop. *Pattern recognition and machine learning*. Springer New York, Secaucus, NJ, USA, 1 edition, 2006. (Cited on page 217.)
- [16] Aaron G Blankenship and Marla B Feller. Mechanisms underlying spontaneous patterned activity in developing neural circuits. *Nature Reviews Neuroscience*, 11(1):18–29, 2009. (Cited on pages 6, 19, and 20.)
- [17] Martin Boerlin, Christian K. Machens, and Sophie Denève. Predictive Coding of Dynamical Variables in Balanced Spiking Networks. *PLoS Computational Biology*, 9(11), 2013. (Cited on page 78.)
- [18] Romain Brette. What Is the Most Realistic Single-Compartment Model of Spike Initiation? *PLoS Computational Biology*, 11(4):1–13, 2015. (Cited on page 26.)
- [19] Rasmus Bro and Henk Kiers. Cross-validation of component models: A critical look at current methods. *Analytical and bioanalytical chemistry*, 5(April):1241–1251, 2008. (Cited on pages 218, 223, 224, and 229.)
- [20] Carlos D Brody, Ranulfo Romo, and Adam Kepecs. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Current Opinion in Neurobiology*, 13(2):204–211, apr 2003. (Cited on page 100.)
- [21] Dean V Buonomano and Wolfgang Maass. State-dependent computations: spatiotemporal processing in cortical networks. *Nature reviews. Neuroscience*, 10(2):113–125, 2009. (Cited on page 78.)
- [22] M. a. Castro-Alamancos and T. Gulati. Neuromodulators Produce Distinct Activated States in Neocortex. *Journal of Neuroscience*, 34(37):12353–12367, 2014. (Cited on page 12.)

- [23] Manuel A. Castro-Alamancos. Absence of Rapid Sensory Adaptation in Neocortex during Information Processing States. *Neuron*, 41(3):455–464, 2004. (Cited on pages [16](#) and [31](#).)
- [24] Raymond Cattell. the scree test for the number of factors. *Multivariate Behavioral Research*, 1(april):245–276, 1966. (Cited on pages [217](#) and [218](#).)
- [25] Elizabeth A. Clement, Alby Richard, Megan Thwaites, Jonathan Ailon, Steven Peters, and Clayton T. Dickson. Cyclic and sleep-like spontaneous alternations of brain state under urethane anaesthesia. *PLoS ONE*, 3(4), 2008. (Cited on pages [13](#) and [79](#).)
- [26] Marlene R Cohen and John H R Maunsell. Attention improves performance primarily by reducing interneuronal correlations. *Nature neuroscience*, 12(12):1594–1600, 2009. (Cited on page [10](#).)
- [27] Albert Compte, Maria V Sanchez-vives, David A McCormick, and Xiao-jing Wang. Cellular and Network Mechanisms of Slow Oscillatory Activity (  $\sim 1$  Hz) and Wave Propagations in a Cortical Network Model. 89(5):2707–2725, 2003. (Cited on page [31](#).)
- [28] Rosa Cossart, Dmitriy Aronov, and Rafael Yuste. Attractor dynamics of network UP states in the neocortex. *Nature*, 423(May):283–288, 2003. (Cited on page [39](#).)
- [29] Lee Cossell, Maria Florencia Iacaruso, Dylan R. Muir, Rachael Houlton, Elie N. Sader, Ho Ko, Sonja B. Hofer, and Thomas D. Mrsic-Flogel. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, 518(7539):399–403, feb 2015. (Cited on pages [183](#) and [203](#).)
- [30] Sylvain Crochet and Carl C H Petersen. Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nature neuroscience*, 9(5):608–610, 2006. (Cited on pages [7](#) and [17](#).)
- [31] C Curto, S Sakata, S Marguet, V Itskov, and K D Harris. A simple model of cortical dynamics explains variability and state-dependence of sensory responses in urethane-anesthetized auditory cortex. *J. Neurosci.*, 29(34):10600–10612, 2009. (Cited on pages [14](#) and [15](#).)
- [32] C Curto, S Sakata, S Marguet, V Itskov, and K D Harris. A simple model of cortical dynamics explains variability and state-dependence of sensory responses in urethane-anesthetized auditory cortex. *J. Neurosci.*, 29(34):10600–10612, 2009. (Cited on page [31](#).)
- [33] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience*. The MIT Press, Cambridge, Massachusetts, mit press edition, 2001. (Cited on pages [98](#) and [148](#).)
- [34] Gustavo Deco and Alexander Thiele. Cholinergic control of cortical network interactions enables feedback-mediated attentional modulation. *European Journal of Neuroscience*, 34(1):146–157, 2011. (Cited on page [11](#).)

- [35] Alain Destexhe. High-conductance state. *Scholarpedia*, 2(11):1341, 2007. (Cited on page [13](#).)
- [36] Giancarlo Diana and Chiara Tommasi. Cross-validation methods in principal component analysis: A comparison. *Statistical Methods & Applications*, 11(1):71–82, feb 2002. (Cited on pages [218](#) and [224](#).)
- [37] Brent Doiron, Ashok Litwin-Kumar, Robert Rosenbaum, Gabriel K Ocker, and Krešimir Josić. The mechanics of state-dependent neural correlations. *Nature Neuroscience*, 19(3):383–393, 2016. (Cited on page [32](#).)
- [38] Rodney Douglas, Christof Koch, M. Mahowald, K. Martin, and H. Suarez. Recurrent excitation in neocortical circuits. *Science*, 269(5226):981–985, 1995. (Cited on pages [102](#), [182](#), [208](#), [251](#), and [252](#).)
- [39] Alexander S Ecker, Philipp Berens, Georgios a Keliris, Matthias Bethge, Nikos K Logothetis, and Andreas S Tolias. Decorrelated Neuronal Firing in Cortical Microcircuits. *Science*, 327(5965):584–587, jan 2010. (Cited on page [40](#).)
- [40] Erika E Fanselow and M A Nicolelis. Behavioral modulation of tactile responses in the rat somatosensory system. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 19(17):7603–7616, 1999. (Cited on page [16](#).)
- [41] Louis Ferre. Selection of components in principal component analysis: A comparison of methods. *Computational Statistics & Data Analysis*, 19(6):669–682, 1995. (Cited on page [237](#).)
- [42] Jozsef Fiser, Pietro Berkes, Gergo Orban, and Mate Lengyel. Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, 14(3):119–130, mar 2010. (Cited on pages [22](#) and [78](#).)
- [43] József Fiser, Chiayu Chiu, and Michael Weliky. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature*, 431(7008):573–578, 2004. (Cited on pages [22](#) and [33](#).)
- [44] U Birte Forstmann and Eric-Jan Wagenmakers. *Model-Based Cognitive Neuroscience: A Conceptual Introduction*, chapter 3, pages 139–156. Springer New York, New York, NY, 2015. (Cited on pages [21](#), [22](#), and [78](#).)
- [45] S Ganguli, D Huh, and H Sompolinsky. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 105(48):18970–18975, 2008. (Cited on page [102](#).)
- [46] Surya Ganguli, James W. Bisley, Jamie D. Roitman, Michael N. Shadlen, Michael E. Goldberg, and Kenneth D. Miller. One-Dimensional Dynamics of Attention and Decision Making in LIP. *Neuron*, 58(1):15–25, 2008. (Cited on pages [33](#), [78](#), [79](#), and [246](#).)
- [47] Crispin Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. 1994. (Cited on pages [146](#) and [147](#).)



## Bibliography

---

- [48] Andrew Gelman. *Bayesian Data Analysis*. Boston, crc press edition, 2014. (Cited on page [236](#).)
- [49] G L Gerstein and B Mandelbrot. Random Walk Models for the Spike Activity of a Single Neuron. *Biophysical journal*, 4(c):41–68, 1964. (Cited on pages [24](#), [25](#), [26](#), and [27](#).)
- [50] Iris Ginzburg and Haim Sompolinsky. Theory of correlations in stochastic neural networks. *Physical Review E*, 50(4):3171–3191, oct 1994. (Cited on page [30](#).)
- [51] Michael Goard and Yang Dan. Basal forebrain activation enhances cortical coding of natural scenes. *Nature Neuroscience*, 12(11):1444–1449, nov 2009. (Cited on page [11](#).)
- [52] Mark S Goldman. Memory without feedback in a neural network. *Neuron*, 61(4):621–34, feb 2009. (Cited on pages [78](#), [100](#), and [102](#).)
- [53] Dan F M Goodman and Romain Brette. The brain simulator. *Frontiers in Neuroscience*, 3(SEP):192–197, 2009. (Cited on pages [106](#) and [146](#).)
- [54] E Guigon, B Dorizzi, Y Burnod, and W Schultz. Neural correlates of learning in the prefrontal cortex of the monkey: a predictive model. *Cerebral Cortex*, 5:135–147, 1995. (Cited on page [26](#).)
- [55] Boris Gutkin and Fleur Zeldenrust. Spike frequency adaptation. *Scholarpedia*, 9(2):30643, 2014. (Cited on page [32](#).)
- [56] Louis Guttman. Some necessary conditions for common factor analysis. *Psychometrika*, 19(2), 1954. (Cited on page [218](#).)
- [57] R H Hahnloser, R Sarpeshkar, M a Mahowald, R J Douglas, and H S Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789):947–951, 2000. (Cited on page [100](#).)
- [58] M. Gartz Hanson, Louise D. Milner, and Lynn T. Landmesser. Spontaneous rhythmic activity in early chick spinal cord influences distinct motor axon pathfinding decisions. *Brain Research Reviews*, 57(1):77–85, 2008. (Cited on page [20](#).)
- [59] Kenneth D Harris and Thomas D Mrsic-Flogel. Cortical connectivity and sensory coding. *Nature*, 503(7474):51–8, 2013. (Cited on pages [2](#), [3](#), and [185](#).)
- [60] Kenneth D. Harris and Gordon M G Shepherd. The neocortical circuit: themes and variations. *Nature Neuroscience*, 18(2):170–181, jan 2015. (Cited on pages [2](#) and [3](#).)
- [61] Kenneth D Harris and Alexander Thiele. Cortical state and attention. *Nature Reviews Neuroscience*, 12(9):509–523, 2011. (Cited on pages [6](#), [7](#), [11](#), [16](#), and [23](#).)
- [62] Christopher D Harvey, Forrest Collman, Daniel A Dombeck, and David W Tank. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature*, 461(7266):941–6, 2009. (Cited on page [9](#).)

- [63] James C. Hayton, David G. Allen, and Vida Scarpello. Factor Retention Decisions in Exploratory Factor Analysis: a Tutorial on Parallel Analysis. *Organizational Research Methods*, 7(2):191–205, 2004. (Cited on page [219](#).)
- [64] Guillaume Hennequin, Tim P. Vogels, and Wulfram Gerstner. Non-normal amplification in random balanced neuronal networks. *Physical Review E*, 86(1):011909, jul 2012. (Cited on pages [98](#), [103](#), [123](#), [127](#), [149](#), [160](#), and [252](#).)
- [65] Guillaume Hennequin, Tim P. Vogels, and Wulfram Gerstner. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron*, 82(6):1394–1406, 2014. (Cited on page [78](#).)
- [66] J Allan Hobson. REM sleep and dreaming: towards a theory of protoconsciousness. *Nature reviews. Neuroscience*, 10(11):803–813, 2009. (Cited on pages [11](#) and [12](#).)
- [67] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of physiology*, 52(1-2):25–71, 1952. (Cited on page [25](#).)
- [68] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554–2558, 1982. (Cited on page [78](#).)
- [69] John L. Horn. A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185, 1965. (Cited on pages [217](#), [218](#), and [242](#).)
- [70] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933. (Cited on page [214](#).)
- [71] Tao Hu and Dmitri B Chklovskii. Reconstruction of sparse circuits using multi-neuronal excitation (RESCUME). *Frontiers in Neuroscience*, 4:1–9, 2010. (Cited on page [207](#).)
- [72] D H Hubel and T N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of physiology*, 160:106–54, jan 1962. (Cited on pages [14](#) and [179](#).)
- [73] D H Hubel and T N Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–43, 1968. (Cited on pages [14](#) and [179](#).)
- [74] Y Ikegaya, G Aaron, R Cossart, Dmitriy Aronov, and I Lampl. Synfire Chains and Cortical Songs: Temporal Modules of Cortical Activity. *Science (New York, NY)*, 559(2004), 2004. (Cited on page [39](#).)
- [75] I.T. Jolliffe. *Principal Component Analysis*. Springer New York, 2002. (Cited on page [218](#).)
- [76] Bjorn M Kampa, Johannes J Letzkus, and Greg J Stuart. Cortical feed-forward networks for binding different streams of sensory information. *Nature Neuroscience*, 9(12):1472–1473, 2006. (Cited on pages [4](#), [5](#), and [71](#).)

- [77] Karl Kandler, Amanda Clause, and Jihyun Noh. Tonotopic reorganization of developing auditory brainstem circuits. *Nature neuroscience*, 12(6):711–7, 2009. (Cited on page [20](#).)
- [78] Matthew T Kaufman, Mark M Churchland, Stephen I Ryu, and Krishna V Shenoy. Cortical activity in the null space: permitting preparation without movement. *Nature neuroscience*, 17(3):440–8, 2014. (Cited on pages [18](#), [76](#), and [78](#).)
- [79] Ian R. Kemp and Birger R. Kaada. The relation of hippocampal theta activity to arousal, attentive behaviour and somato-motor movements in unrestrained cats. *Brain Research*, 95(2-3):323–342, 1975. (Cited on pages [19](#) and [23](#).)
- [80] Tal Kenet, Dmitri Bibitchkov, Misha Tsodyks, Amiram Grinvald, and Amos Arieli. Spontaneously emerging cortical representations of visual attributes. *Nature*, 425(6961):954–956, 2003. (Cited on pages [22](#) and [100](#).)
- [81] T. Kiritani, I. R. Wickersham, H. S. Seung, and G. M. G. Shepherd. Hierarchical Connectivity and Connection-Specific Dynamics in the Corticospinal-Corticostriatal Microcircuit in Mouse Motor Cortex. *Journal of Neuroscience*, 32(14):4992–5001, 2012. (Cited on pages [5](#), [71](#), and [78](#).)
- [82] Ho Ko, Lee Cossell, Chiara Baragli, Jan Antolik, Claudia Clopath, Sonja B. Hofer, and Thomas D. Mrsic-Flogel. The emergence of functional microcircuits in visual cortex. *Nature*, 496(7443):96–100, apr 2013. (Cited on pages [32](#) and [144](#).)
- [83] Walter M. Yamada. Larry F. Abbott, Paul R. Adams, Hagai Agmon-Snir, James M. Bower, Robert E. Burke, Erik de Schutter, Alain Destexhe, Rodney Douglas, Bard Ermentrout, Fabrizio Gabbiani, David Hansel, Michael Hines, Christof Koch, Misha Mahowald, Zachary F. Mainen, Eve Mar. *Methods in Neuronal Modeling, From Ions to Networks*. The MIT Press, Cambridge MA. (Cited on pages [97](#) and [100](#).)
- [84] P E Latham, B J Richmond, S Nirenberg, P G Nelson, and Los Angeles. Intrinsic Dynamics in Neuronal Networks. II. Experiment. *Journal of neurophysiology*, 83(2):808–827, 2000. (Cited on page [31](#).)
- [85] PE Latham and BJ Richmond. Intrinsic dynamics in neuronal networks. I. Theory. *Journal of neurophysiology*, 83(2):808–827, 2000. (Cited on page [31](#).)
- [86] Irvine) Lichman, M (U. California. UCI Machine Learning Repository. *JML*, 1(1):1, 2013. (Cited on page [236](#).)
- [87] Anthony D Lien and Massimo Scanziani. Tuned thalamic excitation is amplified by visual cortical circuits. *Nature neuroscience*, 16(9):1315–23, 2013. (Cited on pages [182](#), [183](#), [209](#), and [252](#).)
- [88] Ashok Litwin-Kumar and Brent Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature neuroscience*, 15(11):1498–505, 2012. (Cited on page [140](#).)

## Bibliography

---

- [89] Artur Luczak, Peter Barthó, and Kenneth D Harris. Spontaneous Events Outline the Realm of Possible Sensory Responses in Neocortical Populations. *Neuron*, 62(3):413–425, 2009. (Cited on pages [15](#), [16](#), [22](#), [23](#), [33](#), [39](#), [73](#), and [74](#).)
- [90] Artur Luczak, Peter Bartho, and Kenneth D. Harris. Gating of Sensory Input by Spontaneous Cortical Activity. *The Journal of Neuroscience*, 33(4):1684–1695, 2013. (Cited on pages [23](#), [73](#), [74](#), [75](#), and [77](#).)
- [91] Artur Luczak, Peter Barthó, Stephan L Marguet, György Buzsáki, and Kenneth D Harris. Sequential structure of neocortical spontaneous activity in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 104(1):347–352, 2007. (Cited on page [39](#).)
- [92] Artur Luczak, Bruce L McNaughton, and Kenneth D Harris. Packet-based communication in the cortex. *Nature reviews. Neuroscience*, 16(12):745–755, 2015. (Cited on pages [23](#) and [74](#).)
- [93] Jakob H Macke, Lars Buesing, Maneesh Sahani, Gatsby Computational, and Neuroscience Unit. Estimating State and Parameters in State Space Models of Spike Trains. In *Advanced State Space Methods for Neural and Clinical Data*, chapter 1, pages 1–26. Cambridge University Press, 2015. (Cited on page [217](#).)
- [94] Zachary F Mainen and Terrence J Sejnowski. Reliability of spike timing in neocortical neurons. *Science (New York, N.Y.)*, 268(5216):1503–1506, 1995. (Cited on page [26](#).)
- [95] Nicolas Maingret, Gabrielle Girardeau, Ralitsa Todorova, Marie Goutierre, and Michaël Zugaro. Hippocampo-cortical coupling mediates memory consolidation during sleep. *Nature Neuroscience*, 19(April):959–964, 2016. (Cited on page [20](#).)
- [96] S. L. Marguet and K. D. Harris. State-Dependent Representation of Amplitude-Modulated Noise Stimuli in Rat Auditory Cortex. *Journal of Neuroscience*, 31(17):6414–6420, 2011. (Cited on pages [15](#), [16](#), and [17](#).)
- [97] David Marr. Simple memory: a theory for archicortex. *Phil. Trans. Royal Soc. London*, 262:23–81., 1(841):23–81, 1971. (Cited on page [20](#).)
- [98] Matthew J. McGinley, Martin Vinck, Jacob Reimer, Renata Batista-Brito, Edward Zagher, Cathryn R. Cadwell, Andreas S. Tolias, Jessica A. Cardin, and David A. McCormick. Waking state: rapid variations modulate neural and behavioral responses. *Neuron*, 87(6):1143–1161, 2015. (Cited on pages [8](#) and [11](#).)
- [99] Jude F. Mitchell, Kristy A. Sundberg, and John H. Reynolds. Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron*, 63(6):879–888, 2009. (Cited on pages [10](#) and [15](#).)
- [100] PP P Mitra and B Pesaran. Analysis of Dynamic Brain Imaging Data. *Biophys J*, 76(2):691–708, 1999. (Cited on page [85](#).)

## Bibliography

---

- [101] Gabriela Mochol, Ainhoa Hermoso-Mendizabal, Shuzo Sakata, Kenneth D Harris, and Jaime de la Rocha. Stochastic transitions into silence cause noise correlations in cortical circuits. *Proceedings of the National Academy of Sciences*, 112(11):201410509, 2015. (Cited on page [31](#).)
- [102] Alik Mokeichev, Michael Okun, Omri Barak, Yonatan Katz, Ohad Ben-Shahar, and Ilan Lampl. Stochastic Emergence of Repeating Cortical Motifs in Spontaneous Membrane Potential Fluctuations In Vivo. *Neuron*, 53(3):413–425, 2007. (Cited on page [39](#).)
- [103] J Moran and R Desimone. Selective attention gates visual processing in the extrastriate cortex. *Science (New York, N.Y.)*, 229(4715):782–784, 1985. (Cited on page [10](#).)
- [104] Nicolás A Morgenstern, Jacques Bourg, and Leopoldo Petreanu. Multilaminar networks of cortical neurons integrate common inputs from sensory thalamus. *Nature Neuroscience*, 19(8):1034–1040, jul 2016. (Cited on pages [2](#), [184](#), [202](#), [203](#), [204](#), [205](#), and [252](#).)
- [105] Brendan K. Murphy and Kenneth D. Miller. Balanced Amplification: A New Mechanism of Selective Amplification of Neural Activity Patterns. *Neuron*, 61(4):635–648, 2009. (Cited on pages [33](#), [98](#), [99](#), [100](#), [101](#), [102](#), [104](#), [105](#), [109](#), [110](#), [135](#), [136](#), [154](#), [182](#), [246](#), [251](#), and [252](#).)
- [106] John D Murray, Alberto Bernacchia, David J Freedman, Ranulfo Romo, Jonathan D Wallis, Xinying Cai, Camillo Padoa-Schioppa, Tatiana Pasternak, Hyojung Seo, Daeyeol Lee, and Xiao-Jing Wang. A hierarchy of intrinsic timescales across primate cortex. *Nature neuroscience*, 17(12):1661–3, 2014. (Cited on page [18](#).)
- [107] Cristopher M. Niell and Michael P. Stryker. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron*, 65(4):472–479, 2010. (Cited on pages [9](#) and [14](#).)
- [108] Cian O’Donnell and Mark C W van Rossum. Spontaneous action potentials and neural coding in unmyelinated axons. *Neural computation*, 27(4):801–18, 2015. (Cited on page [25](#).)
- [109] Michael Okun, Nicholas A Steinmetz, Lee Cossell, M Florencia Iacaruso, Ho Ko, Péter Barthó, Tirin Moore, Sonja B Hofer, Thomas D Mrsic-Flogel, Matteo Carandini, and Kenneth D Harris. Diverse coupling of neurons to populations in sensory cortex. *Nature*, 521(7553):511–515, 2015. (Cited on pages [39](#), [51](#), [54](#), [58](#), [72](#), and [73](#).)
- [110] Shawn R. Olsen, Dante S. Bortone, Hillel Adesnik, and Massimo Scanziani. Gain control by layer six in cortical circuits of vision. *Nature*, 483(7387):47–52, 2012. (Cited on pages [2](#) and [5](#).)
- [111] Marius Pachitariu, Dmitry R Lyamzin, Maneesh Sahani, and Nicholas a Lesica. State-dependent population coding in primary auditory cortex. *Journal of neuroscience*, 35(5):2058–73, 2015. (Cited on pages [14](#), [15](#), and [16](#).)

## Bibliography

---

- [112] Karl Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(1):559–572, 1901. (Cited on page [214](#).)
- [113] Donald B. Percival and Andrew T. Walden. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge University Press, 1993. (Cited on page [85](#).)
- [114] Pedro R. Peres-Neto, Donald A. Jackson, and Keith M. Somers. How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis*, 49(4):974–997, 2005. (Cited on pages [235](#), [236](#), [238](#), [239](#), and [242](#).)
- [115] Leopoldo Petreanu, Daniel Huber, Aleksander Sobczyk, and Karel Svoboda. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature neuroscience*, 10(5):663–8, 2007. (Cited on pages [185](#), [186](#), and [189](#).)
- [116] Leopoldo Petreanu, Tianyi Mao, Scott M Sternson, and K Svoboda. The subcellular organization of neocortical excitatory connections. *Nature*, 457(7233):1142–5, 2009. (Cited on page [187](#).)
- [117] Dietmar Plenz and Tara C. Thiagarajan. The organizing principles of neuronal avalanches: cell assemblies in the cortex? *Trends in Neurosciences*, 30(3):101–110, 2007. (Cited on page [39](#).)
- [118] Pierre Olivier Polack, Jonathan Friedman, and Peyman Golshani. Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nature neuroscience*, 16(9):1331–9, 2013. (Cited on page [12](#).)
- [119] J F a Poulet and C C H Petersen. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature*, 454(7206):881–885, 2008. (Cited on page [9](#).)
- [120] Yifat Prut, Eilon Vaadia, Hagai Bergman, Iris Haalman, Hamutal Slovin, Moshe Abeles, Eilon Vaadia, Hagai Bergman, and Iris Haalman. Spatiotemporal Structure of Cortical Activity : Properties and Behavioral Relevance. pages 2857–2874, 1998. (Cited on page [39](#).)
- [121] E. Racca and A. Porporato. Langevin equations from time series. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 71(2):18–20, 2005. (Cited on page [109](#).)
- [122] Kanaka Rajan and L. F. Abbott. Eigenvalue Spectra of Random Matrices for Neural Networks. *Physical Review Letters*, 97(18):188104, 2006. (Cited on pages [127](#), [158](#), and [160](#).)
- [123] A. Renart, J. de la Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris. The Asynchronous State in Cortical Circuits. *Science*, 327(5965):587–590, 2010. (Cited on pages [27](#), [30](#), [31](#), [32](#), [40](#), [68](#), [79](#), [80](#), [106](#), [107](#), [123](#), [125](#), [146](#), [160](#), [247](#), and [251](#).)

## Bibliography

---

- [124] C. Rock and A. j. Apicella. Callosal Projections Drive Neuronal-Specific Responses in the Mouse Auditory Cortex. *Journal of Neuroscience*, 35(17):6703–6713, 2015. (Cited on pages [71](#) and [72](#).)
- [125] R W Rodieck, N Y Kiang, and G L Gerstein. Some quantitative methods for the study of spontaneous activity of single neurons. *Biophys J*, 2:351–368, 1962. (Cited on page [25](#).)
- [126] Robert Rosenbaum and Brent Doiron. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Physical Review X*, 4(2):1–9, 2014. (Cited on page [144](#).)
- [127] Robert Rosenbaum, Matthew A Smith, Adam Kohn, Jonathan E Rubin, and Brent Doiron. The spatial structure of correlated neuronal variability. *Nature Neuroscience*, (October):1–35, 2016. (Cited on pages [249](#), [250](#), [251](#), and [252](#).)
- [128] Alex Roxin and Anders Ledberg. Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Computational Biology*, 4(3), 2008. (Cited on pages [100](#) and [113](#).)
- [129] Nava Rubin Rubén Moreno-Bote, John Rinzel. Noise-Induced Alternations in an Attractor Network Model of Perceptual Bistability. *Journal of Neurophysiology*, 98(3):1125–1139, 2007. (Cited on pages [100](#) and [113](#).)
- [130] Sepideh Sadaghiani, Guido Hesselmann, and Andreas Kleinschmidt. Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. *The Journal of neuroscience*, 29(42):13410–13417, 2009. (Cited on pages [22](#) and [78](#).)
- [131] M V Sanchez-Vives and D a McCormick. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature Neuroscience*, 3(10):1027–1034, 2000. (Cited on pages [7](#), [12](#), and [145](#).)
- [132] Michael T. Schaub, Yazan N. Billeh, Costas A. Anastassiou, Christof Koch, and Mauricio Barahona. Emergence of Slow-Switching Assemblies in Structured Neuronal Networks. *PLoS Computational Biology*, 11(7):1–28, 2015. (Cited on pages [114](#) and [140](#).)
- [133] David M. Schneider, Anders Nelson, and Richard Mooney. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517):189–94, 2014. (Cited on page [14](#).)
- [134] Odelia Schwartz, Jonathan W Pillow, Nicole C Rust, and Eero P Simoncelli. Spike-triggered neural characterization. *Journal of vision*, 6(4):484–507, 2006. (Cited on pages [225](#), [230](#), and [231](#).)
- [135] William Beecher Scoville and Brenda Milner. Memory and the Medial Temporal Lobe: Patient HM. *J Neuropsychiatry Clin Neurosci*, 253(5026):1380–1386, 1957. (Cited on page [20](#).)



## Bibliography

---

- [136] Michael N. Shadlen and William T. Newsome. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4(4):569–579, 1994. (Cited on pages [27](#) and [30](#).)
- [137] Michael N Shadlen and William T Newsome. The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding. *The Journal of Neuroscience*, 18(10):3870–3896, 1998. (Cited on pages [27](#) and [32](#).)
- [138] Anton Sirota, Jozsef Csicsvari, Derek Buhl, and G Buzsáki. Communication between neocortex and hippocampus during sleep in rodents. *Proceedings of the National Academy of Sciences of the United States of America*, 100(4):2065–2069, 2003. (Cited on page [20](#).)
- [139] W. R. Softky and C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 13(1):334–50, 1993. (Cited on pages [25](#) and [27](#).)
- [140] Nicholas C Spitzer. Electrical activity in early neuronal development. *Nature*, 444(7120):707–712, 2006. (Cited on page [19](#).)
- [141] David Sussillo and L F Abbott. Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4):544–57, aug 2009. (Cited on page [78](#).)
- [142] Andrew Y Y Tan, Yuzhi Chen, Benjamin Scholl, Eyal Seidemann, and Nicholas J Priebe. Sensory stimulation shifts visual cortex from synchronous to asynchronous states. *Nature*, 509(7499):226–9, 2014. (Cited on page [7](#).)
- [143] Bosiljka Tasic, Vilas Menon, Thuc Nghi Nguyen, Tae Kyung Kim, Tim Jarsky, Zizhen Yao, Boaz Levi, Lucas T Gray, Staci A Sorensen, Tim Dolbeare, Darren Bertagnolli, Jeff Goldy, Nadiya Shapovalova, Sheana Parry, Changkyu Lee, Kimberley Smith, Amy Bernard, Linda Madisen, Susan M Sunkin, Michael Hawrylycz, Christof Koch, and Hongkui Zeng. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience*, 19(2):335–346, jan 2016. (Cited on page [2](#).)
- [144] Tom Tetzlaff, Moritz Helias, Gaute T. Einevoll, and Markus Diesmann. Decorrelation of Neural-Network Activity by Inhibitory Feedback. *PLoS Computational Biology*, 8(8), 2012. (Cited on page [31](#).)
- [145] D. J. Tolhurst, J. A. Movshon, and A. F. Dean. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, 23(8):775–785, 1983. (Cited on page [19](#).)
- [146] Todd W. Troyer and Kenneth D. Miller. Physiological Gain Leads to High ISI Variability in a Simple Model of a Cortical Regular Spiking Cell, 1997. (Cited on page [27](#).)
- [147] M Tsodyks, T Kenet, and A Arieli. Linking Spontaneous Activity of Single Cortical Neurons and The Underlying Functionl Architecture. *Science*, 286(1999):1943, 1999. (Cited on pages [33](#) and [100](#).)



## Bibliography

---

- [148] Misha Tsodyks and Terrence J Sejnowski. Rapid state switching in balanced cortical network models. *Network: Computation in Neural Systems*, 6:111–124, 1995. (Cited on page [27](#).)
- [149] C van Vreeswijk and H Sompolinsky. Chaotic Balanced State in a Model of Cortical Circuits. *Neural Computation*, 10(6):1321–1371, aug 1998. (Cited on pages [24](#), [27](#), [28](#), [31](#), and [106](#).)
- [150] C H Vanderwolf. Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalography and clinical neurophysiology*, 26(4):407–418, 1969. (Cited on pages [8](#) and [19](#).)
- [151] C. H. Vanderwolf. Cerebral Activity and Behavior: Control by Central Cholinergic and Serotonergic Systems. *International Review of Neurobiology*, 30(C):225–340, 1988. (Cited on page [11](#).)
- [152] Wayne F. Velicer. Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 1976. (Cited on pages [218](#) and [221](#).)
- [153] Wayne F Velicer, Cheryl A Eaton, and Joseph L Fava. Construct Explication through Factor or Component Analysis: A Review and Evaluation of Alternative Procedures for Determining the Number of Factors or Components. In *Problems and solutions in human assessment*, chapter 2, pages 47–71. Kluwer, Boston, 2000. (Cited on pages [222](#) and [242](#).)
- [154] Martin Vinck, Jeroen J Bos, Laura A Van Mourik-Donga, Krista T Oplaat, Gerbrand A Klein, Jadin C Jackson, Luc J Gentet, and Cyriel M A Pennartz. Cell-Type and State-Dependent Synchronization among Rodent Somatosensory, Visual, Perirhinal Cortex, and Hippocampus CA1. *Frontiers in systems neuroscience*, 9(January):187, 2015. (Cited on page [58](#).)
- [155] Xiao-Jing Wang. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36(5):955–68, dec 2002. (Cited on pages [100](#), [113](#), [139](#), and [246](#).)
- [156] L. M. Ward. Attention. *Scholarpedia*, 3(10):1538, 2008. (Cited on page [9](#).)
- [157] Michael C Wiest and Miguel A L Nicolelis. Behavioral detection of tactile stimuli during 7–12 Hz cortical oscillations in awake rats. *Nature Neuroscience*, 6(9):913–914, 2003. (Cited on pages [17](#) and [19](#).)
- [158] Ryan C Williamson, Benjamin R Cowley, Ashok Litwin-Kumar, Brent Doiron, Adam Kohn, Matthew A Smith, and Byron M Yu. Scaling Properties of Dimensionality Reduction for Neural Populations and Network Models. *PLoS computational biology*, 12(12):e1005141, 2016. (Cited on pages [108](#) and [143](#).)
- [159] Klaus Wimmer, Albert Compte, Alex Roxin, Diogo Peixoto, Alfonso Renart, and Jaime de la Rocha. Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT. *Nature communications*, 6(February):6177, 2015. (Cited on page [32](#).)

- [160] K.-F. Wong. A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. *Journal of Neuroscience*, 26(4):1314–1328, 2006. (Cited on pages [100](#), [113](#), and [139](#).)
- [161] F Wörgötter, K Suder, Y Zhao, N Kerscher, U T Eysel, and K Funke. State-dependent receptive-field restructuring in the visual cortex. *Nature*, 396(6707):165–8, 1998. (Cited on page [14](#).)
- [162] Robert H. Wurtz. Visual receptive fields of striate cortex neurons in awake monkeys. *Journal of neurophysiology*, 32(5):727–742, 1969. (Cited on page [14](#).)
- [163] Min Xu, Shinjae Chung, Siyu Zhang, Peng Zhong, Chenyan Ma, Wei-Cheng Chang, Brandon Weissbourd, Noriaki Sakai, Liqun Luo, Seiji Nishino, and Yang Dan. Basal forebrain circuit for sleep-wake control. *Nature Neuroscience*, 18(11):1641–1647, 2015. (Cited on page [11](#).)
- [164] Yumiko Yoshimura, Jami L M Dantzker, and Edward M Callaway. Excitatory cortical neurons from fine-scale functional networks. *Nature*, 5(February):2005–2005, 2005. (Cited on pages [4](#), [71](#), and [184](#).)
- [165] Bm Yu, Jp Cunningham, G Santhanam, Si Ryu, Kv Shenoy, and M Sahani. Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity (vol 102, pg 614, 2009). *Journal of Neurophysiology*, 102(April 2009):614–635, 2009. (Cited on page [217](#).)
- [166] Edward Zagher, Xinxin Ge, and David A. McCormick. Competing Neural Ensembles in Motor Cortex Gate Goal-Directed Motor Output. *Neuron*, 88(3):565–577, 2015. (Cited on pages [7](#), [9](#), [72](#), [113](#), [139](#), [140](#), and [141](#).)
- [167] Gil Ziv, Barry W. Connors, and Yael Amitai. Efficacy of thalamocortical and intracortical synaptic connections: Quanta, innervation, and reliability. *Neuron*, 23(2):385–397, 1999. (Cited on page [199](#).)
- [168] Ehud Zohary, Michael N. Shadlen, and William T. Newsome. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370(6485):140–143, jul 1994. (Cited on page [30](#).)

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal  
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

**[www.itqb.unl.pt](http://www.itqb.unl.pt)**

